MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

②

AD-A187 572

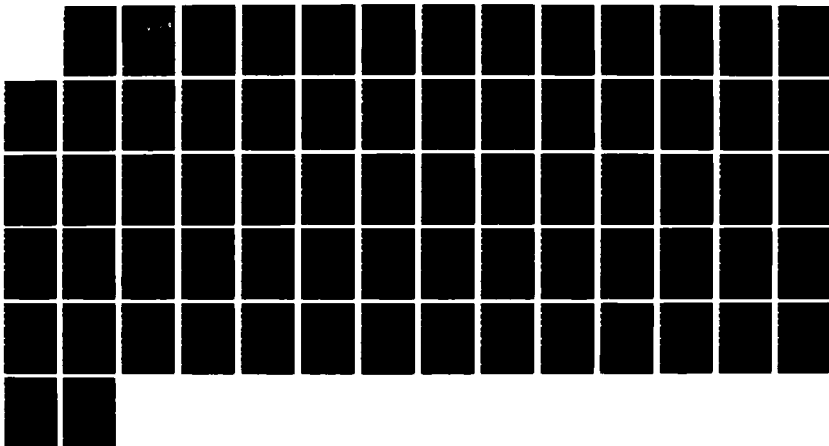# NAVAL POSTGRADUATE SCHOOL
Monterey, California

DTIC
S ELECTE
DEC 1 6 1987
D

THESIS

A COMPARISON OF PROJECTION METHODS IN THE
FORECASTING OF OVERHEAD COSTS FOR SEVEN
GOVERNMENT AEROSPACE CONTRACTORS

by

Christopher P. Schnedar

September 1987

Thesis Advisor: D. C. Boger

87 12 9 062

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b RESTRICTIVE MARKINGS			
2a SECURITY CLASSIFICATION AUTHORITY		3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.			
2b DECLASSIFICATION/DOWNGRADING SCHEDULE					
4 PERFORMING ORGANIZATION REPORT NUMBER(S)		5 MONITORING ORGANIZATION REPORT NUMBER(S)			
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School	6b OFFICE SYMBOL (If applicable) 55	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School			
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000		7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000			
8a NAME OF FUNDING/SPONSORING ORGANIZATION	8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER			
8c ADDRESS (City, State and ZIP Code)		10 SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification) A Comparison of Projection Methods in the Forecasting of Overhead Costs for Seven Government Aerospace Contractors

12 PERSONAL AUTHOR(S) SCHNEDAR, Christopher P.

13a TYPE OF REPORT Master's Thesis	13b TIME COVERED FROM ___ TO ___	14 DATE OF REPORT (Year, Month, Day) 1987 September	15 PAGE COUNT 69

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) Box-Jenkins, Regression, Autocorrelation
FIELD	GROUP	SUB GROUP	

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

    This thesis compares three types of models developed to predict overhead costs for seven government aerospace contractors. The methodologies utilized to develop the models include generalized least squares, univariate Box-Jenkins, and multivariate Box-Jenkins procedures. The results of those models are compared using three measures of effectiveness: correlation coefficient between actual and predicted values, root mean squared error divided by the mean of the actuals, and mean absolute percentage error (in percent). As was expected, the univariate Box-Jenkins method produced short term forecasts which were superior to those of the least squares regression models. However, the regression forecasts were highly accurate and were considerably less expensive to obtain. Only one multivariate Box-

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a NAME OF RESPONSIBLE INDIVIDUAL Prof. D. C. Boger	22b TELEPHONE (Include Area Code) 408-646-2607	22c OFFICE SYMBOL 55Bo

DD FORM 1473, 84 MAR   83 APR edition may be used until exhausted   SECURITY CLASSIFICATION OF THIS PAGE
                         All other editions are obsolete

Block 19 Abstract Continued

Jenkins model could be developed. The results of this model were marginally superior to the related regression model and significantly inferior to the univariate Box-Jenkins model for the same contractor.

A Comparison of Projection Methods in the Forecasting
Of Overhead Costs for Seven Government Aerospace
Contractors

by

Christopher P. Schnedar
Lieutenant Commander, United States Navy
B.B.A.S., University of New Mexico, 1977

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1987

Author: _____
Christopher P. Schnedar

Approved by: _____
D. C. Boger, Thesis Advisor

_____
D. R. Barr, Second Reader

_____
Peter Purdue, Chairman,
Department of Operations Research

_____
James M. Fremgen,
Acting Dean of Information and Policy Sciences

3

ABSTRACT

This thesis compares three types of models developed to predict overhead costs for seven government aerospace contractors. The methodologies utilized to develop the models include generalized least squares, univariate Box-Jenkins, and multivariate Box-Jenkins procedures. The results of those models are compared using three measures of effectiveness: correlation coefficient between actual and predicted values, root mean squared error divided by the mean of the actuals, and mean absolute percentage error (in percent). As was expected, the univariate Box-Jenkins method produced short term forecasts which were superior to those of the least squares regression models. However, the regression forecasts were highly accurate and were considerably less expensive to obtain. Only one multivariate Box-Jenkins model could be developed. The results of this model were marginally superior to the related regression model and significantly inferior to the univariate Box-Jenkins model for the same contractor.

4

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Overhead costs often constitute a large portion of total product costs in many industries. In the case of government aerospace contractors, overhead costs may comprise as much as fifty percent of the contract cost negotiated by the Federal government. Historically, overhead costs have been predicted by applying estimated overhead rates to estimated labor hours in several categories of operations. These category totals are summed to arrive at a total overhead cost for a particular product. This procedure of estimating overhead is greatly affected by changes in the level of operations. Fluctuating production rates often lead to overhead rates which display a lagged relationship between the applied rates and actual overhead costs. Consequently, this particular method of predicting overhead costs produces estimates which are inadequate.

Alternative methods for predicting overhead have been proposed. In many cases, these methods attempt to establish a direct relationship between costs and factors related to direct production. Thus the need for reliance on estimated overhead rates is eliminated. Various least squares regression models have been proposed in attempts to establish the desired direct relationship. One such model was developed with simplicity in application as a strong consideration [Ref. 1: p.7]. This allowed users of varying degrees of statistical familiarity to apply the model in actual work conditions. This model intentionally utilized the minimum number of explanatory variables necessary to achieve accurate results. Several applicable independent variables were considered and observations of direct personnel demonstrated the strongest relationship with overhead costs [Ref. 1: p. 20]. As with any economic trend data, autocorrelation must be suspected. The use of quarterly observations in the model required testing for first order AR(1) and fourth order AR(4) autoregressive processes. Appropriate tests and corrections were incorporated into the model to ensure the absence of bias in the standard errors of the coefficients and in the R-squared statistic [Ref. 2: p.283]. The results of this model are included for comparison purposes.

A second alternative to the use of estimated overhead rates is the Box-Jenkins method of forecasting. This method, which is designed to produce highly accurate short term forecasts, allows for a wide range of possible models to apply to a particular

8

economic series [Ref. 3: p.11]. The degree of statistical sophistication required to apply this method is far greater than that required for the regression model mentioned above. In addition, the use of the Box-Jenkins forecasting method requires far greater computer resources during the three stages of identification, estimation, and forecasting than most regression models. Consequently, the Box-Jenkins method has not been applied as a forecasting tool in many cases in which it would be a logical alternative.

The Box-Jenkins transfer function (multivariate Box-Jenkins method) utilizes deviations from appropriate means of an input (X) and of an output (Y) to establish a relationship from which forecasts can be made. The connecting tool between these two series is a linear differential equation. This extremely complicated forecasting tool requires extensive expertise on the part of the statistician. Once again, this highly effective forecasting tool has been underutilized.

This paper attempts to develop usable least squares regression, Box-Jenkins, and Box-Jenkins transfer function forecasting models for seven government aerospace contractors. The effectiveness of each model is measured by withholding the last four data observations during the model development phase and using the model to forecast the withheld observations. The deviation of the actuals from the predicted values have been indicated with three measures of effectiveness: correlation coefficient between actual and predicted values, root mean squared error divided by the mean of the actuals, and mean absolute percentage error (in percent.) The development and results of each model are included for comparison purposes.

II. DATA

The data have been supplied by seven government aerospace contractors. To maintain confidentiality, all references to specific contractors will be with the labels A through G. A specific reporting format was utilized by each contractor during data collection. Quarterly data spanning the period beginning in the first quarter of 1979 and continuing through the third quarter of 1986 was requested from each of the seven contractors. Usable data were obtained from each contractor for large portions of the requested period. However, only one contractor was able to supply all thirty-one observations.

In the reporting format, overhead costs were composed of costs from three major categories which had been further refined into six subcategories. The first major category, labor related costs, was composed of two subcategories: indirect salaries and fringe benefits. The second major category was facility costs. The third major category is composed of three subcategories: electronic data processing costs, independent research and development and bid and proposal costs, and all other overhead costs.

All components of overhead costs were converted to constant 1984 fourth quarter dollars. The labor related cost categories adjustment was accomplished through the application of the Bureau of Labor Statistics SIC 372 price index for production worker average hourly wages for the aircraft and parts industry. In the case of this index, monthly indices were averaged to produce quarterly indices. Gross National Product Deflater indices published by the Bureau of Economic Analysis were applied to the two remaining major categories. Facility costs were adjusted with the GNPD gross private domestic fixed nonresidential investment index. The GNPD personal consumption services index was utilized to adjust the final major category. As with all indices, those used were imperfect. They were chosen because they provide the best adjustments for inflation among all readily available and relevant indices.

Data pertaining to direct production were obtained from each contractor. The only direct production data set utilized in the analysis was direct labor personnel. This category of data did not require adjustment for inflation.

III. MODEL DEVELOPMENT

A. THE REGRESSION MODEL

The model utilized in this project was designed at the request of the Naval Air Systems Command as part of its contractor overhead tracking project. The model capitalized on the explanatory and prediction properties of least squares theory. Individual regression models were developed for each contractor in an effort to accurately predict future overhead costs. Simplicity in application has been a key factor in the model's development.

The application of least squares to economic trend data almost immediately implies autocorrelation in the error terms of the regression. With the presence of autocorrelation, the estimates of the coefficients are unbiased and consistent. However, they are not efficient. The estimate for the variance of the coefficients are biased. Positively autocorrelated errors produce a coefficient variance which is underestimated because of a downward bias in the estimate of the variance. The downward bias will produce a confidence interval which is narrower than it should be for each coefficient. For this reason, tests of the null hypothesis that the coefficient is equal to 0 will be rejected in instances in which it should be accepted. Likewise, autocorrelated errors will cause exaggerated R^2 and F statistics when an ordinary least squares model is applied.

The effects of autocorrelation can be eliminated through the application of generalized least squares (GLS) procedures. The application of GLS to autocorrelated data produces estimators of the coefficients which possess the properties of maximum likelihood estimators. Therefore, the GLS estimators of the regression coefficients and the variances of these coefficients will be unbiased, consistent, and asymptotically efficient. This will lead to more reliable estimates of R^2 and F. [Ref. 4: pp. 302-311]

This project examined quarterly data. For this reason first order AR(1) and fourth order AR(4) autoregressive processes are suspected. AR(1) processes are detectable with a Durbin Watson Test. In those cases where AR(1) is present, the data has been transformed in the following manner:

$$Z_1^* = Z_1(1 - \hat{\rho}_1^2)^{1/2}$$
$$Z_t^* = Z_t - \hat{\rho}_1 Z_{t-1}. \qquad \text{(eqn 3.1)}$$

11

$$(t = 2,3,4,...,T).$$

where

$$\hat{\rho}_1 = \frac{\displaystyle\sum_{t=2}^{T} e_t e_{t-1}}{\displaystyle\sum_{t=2}^{T-1} e_t^2} \qquad (t = 2,3,...,T) \qquad \text{(eqn 3.2)}$$

where

$$e_t = \text{residuals from the OLS regression.}$$

In equation 3.2, $\hat{\rho}_1$ defined in this manner is the two stage Prais-Winsten estimator derived by Park and Mitchell [Ref. 5]. The AR(4) process detected in the model, a special form of the general AR(4) process, is written as

$$\varepsilon_t = \hat{\rho}_4 \varepsilon_{t-4} + \upsilon_t. \qquad \text{(eqn 3.3)}$$

The general form for AR(4) processes is

$$\varepsilon_t = \hat{\rho}_1 \varepsilon_{t-1} + \hat{\rho}_2 \varepsilon_{t-2} + \hat{\rho}_3 \varepsilon_{t-3} + \hat{\rho}_4 \varepsilon_{t-4} + \upsilon_t.$$

$$\text{(eqn 3.4)}$$

In both cases

$$\upsilon_t \sim N(0,\sigma^2).$$

In this model, the effects of the three prior quarters are assumed to be negligible while the effect of the quarter one year previous is considered significant. This special form of AR(4) process is detectable with a Wallis Test [Ref. 6]. If AR(4) processes were detected, the data sets were transformed in the following manner

$$Z_t^* = Z_t(1 - \hat{\rho}_4^2)^{1/2}$$
$$(t = 1,2,3,4),$$

12

$$Z_t^* = Z_t - \widehat{\rho}_4 Z_{t-1} . \qquad \text{(eqn 3.6)}$$
$$(t = 5,6,7,...,T).$$

where

$$\widehat{\rho}_4 = \frac{T^2(1 - .5d_4) + K^2}{T^2 - K^2} . \qquad \text{(eqn 3.7)}$$

In equation 3.7, T is equal to the number of observations, K is the number of parameters to be estimated, and d_4 is the Wallis Test Statistic written as

$$d_4 = \frac{\displaystyle\sum_{t=5}^{T} (e_t - e_{t-4})^2}{\displaystyle\sum_{t=1}^{T} e_t^2} \qquad (t = 1,2,...,T) \qquad \text{(eqn 3.8)}$$

where

e_t = residuals from the OLS regression model.

This particular estimator, derived by Theil and Nagar [Ref. 7: p. 287], was found to be the most efficient among nine alternative estimaters applied directly to these data sets. [Ref. 8: p. 49]

After each OLS model was computed, the Wallis and the Durbin Watson Statistics were examined for the presence of AR(4) or AR(1) processes. Appropriate transformations were made and the models were reestimated. At the conclusion of the transformation and reestimation phase, all traces of autocorrelation had been removed from the GLS models. In each case, the residuals were examined and all tests for normality were accepted.

The following procedure was applied to each of the seven contractors. A detailed presentation of the results will be made for contractor A. The results from the remaining contractors will be made in a summarized fashion with appropriate

13

comments. Direct comparison of the regression models between two contractors does not indicate relative efficiency. Aside from the organizational differences within each firm, each of these contractors specializes in a particular branch of the aerospace industry. Examples include fixed wing aircraft and aircraft engine producers. For these reasons, comparison of the models between contractors is inappropriate.

Contractor A supplied data spanning twenty-six quarters. Therefore, all three models: regression, Box-Jenkins, and Box-Jenkins transfer function, are based on twenty-two observations. The remaining four observations are withheld during the model development stage and are used for comparison purposes with predicted values during the forecasting stage. The results of the OLS and GLS models are presented in Table 1. The adjusted R^2 value of .1828 and the F value of 5.93 indicate that the OLS model is poor. The Wallis Statistic of .5801 indicates the presense of the special form of AR(4) suspected in the data. The appropriate transformation to correct for AR(4) was made and the model improved significantly: adjusted $R^2 = .8718$ and $F = 150.6569$. At this point, the residuals indicated the effects of first order autocorrelation. A second transformation was made and the model improved slightly: adjusted $R^2 = .8787$ and $F = 160.3358$. Both forms of autocorrelation had been completely removed at the completion of the second transformation.

TABLE 1

REGRESSION MODELS FOR CONTRACTOR A

	OLS	GLS AR(4)	GLS AR(1)
Adjusted R^2	.183	.872	.879
F Statistic	5.921	150.657	160.336
Intercept:	54778.277	11487.623	3047.305
Standard Error:	68215.152	5924.463	2438.277
Slope:	13.960	15.701	15.946
Standard Error:	5.737	1.279	1.259
Durbin Watson Statistic:	1.618	.644	1.991
Wallis Statistic:	.580	1.351	1.432
Estimate of ρ_1	.187	.694	.080
Estimate of ρ_4	.723	.334	.294

The regression models for the remaining six contractors are depicted in Table 2. A single transformation was required to eliminate all indications of autocorrelation in the cases of contractors B and C. The required transformation removed the effects of an AR(4) process. The data supplied by Contractor E was free of the effects of autocorrelation and the model listed in Table 2 is the OLS model developed for this contractor. The remaining contractor data sets were transformed twice to achieve the desired residual characteristics. All of the models derived during this stage possess good forecasting capabilities when utilized to predict the four data points which were withheld. This will be discussed in the next chapter.

B. BOX-JENKINS METHOD

The Box-Jenkins method of forecasting is comprised of three stages: identification, estimation, and forecasting [Ref. 9: p. 19]. In the identification stage, the series is often differenced to achieve stationarity about a mean (usually 0). During the development of these models, differencing of the order of one period (regular differencing) or four periods (seasonal differencing) was considered. As is standard with Box-Jenkins methodology, no more than two differencing corrections were required for any model [Ref. 10: p. 125]. With one exception, a stationary series was achieved through the application of regular, seasonal, or a combination of both types of differencing for each of the contractors. Contractor D data were already stationary and did not require differencing. The autocorrelation (ACF) and partial autocorrelation(PACF) functions of this stationary series are analyzed to determine the horizontal subpatterns within the series. The stationary trend can be specified as a linear combination of past series values (autoregressive terms), a linear combination of past random errors (moving average terms), or a combination of both. The determination of the appropriate number and specific lag of the autoregressive and moving average terms is made during the analysis of the ACF and PACF. Spikes on the ACF accompanied by trends on the PACF resembling exponential decay for the same lag indicate the appropriateness of a moving average parameter at that lag. Likewise, spikes on the PACF accompanied by an exponentially decaying trend on the ACF signal the need to include an autoregressive term at that lag. Spikes on the ACF and PACF are taken to be correlation values for a given lag which are statistically different from 0. Once the character of the trend is identified, the estimation phase is

15

TABLE 2
SUMMARY OF CONTRACTOR REGRESSION MODELS

REGRESSION MODEL for CONTRACTOR B
Model: OVRHD(B) = a + b PERSONNEL(B)

Adjusted R^2	.927
F Statistic	243.881
Intercept:	11978.143
Standard Error:	5952.942
Slope:	15.215
Standard Error:	.974

REGRESSION MODEL for CONTRACTOR C
Model: OVRHD(C) = a + b PERSONNEL(C)

Adjusted R^2	.840
F Statistic	106.326
Intercept:	327.038
Standard Error:	3911.146
Slope:	8.702
Standard Error:	.844

REGRESSION MODEL for CONTRACTOR D
Model: OVRHD(D) = a + b PERSONNEL(D)

Adjusted R^2	.726
F Statistic	64.448
Intercept:	13864.996
Standard Error:	3084.697
Slope:	7.999
Standard Error:	.996

entered to determine the parameter values. A first order autoregressive series is expressed as follows

$$Z_t = \theta_0 + \varphi Z_{t-1} + A_t, \qquad \text{(eqn 3.9)}$$

16

TABLE 2
SUMMARY OF CONTRACTOR REGRESSION MODELS (CONT'D.)

REGRESSION MODEL for CONTRACTOR E

Model: OVRHD(E) = a + b PERSONNEL(E)

Adjusted R^2	.527
F Statistic	26.590
Intercept:	4976.203
Standard Error:	18568.690
Slope:	14.949
Standard Error:	2.899

REGRESSION MODEL for CONTRACTOR F

Model: OVRHD(F) = a + b PERSONNEL(F)

Adjusted R^2	.707
F Statistic	63.739
Intercept:	884.075
Standard Error:	1083.769
Slope:	13.012
Standard Error:	1.629

REGRESSION MODEL for CONTRACTOR G

Model: OVRHD(G) = a + b PERSONNEL(G)

Adjusted R^2	.857
F Statistic	133.155
Intercept:	4716.544
Standard Error:	3020.132
Slope:	20.545
Standard Error:	1.780

where

θ_0 = series mean

φ_1 = weighting of the previous period value

A_t = white noise $\sim N(0, \sigma_A^2)$.

An alternative form of the model in terms of deviation from the series mean is

$$Z_t^* = \varphi Z_{t-1} + A_t \qquad \text{(eqn 3.10)}$$

where

$$Z_t^* = Z_t - \theta_0.$$

This expression can be expanded to include any number of past series values and is written as follows:

$$Z_t^* = \varphi_1 Z_{t-1}^* + \varphi_2 Z_{t-2}^* + \dots + \varphi_p Z_{t-p}^* + A_t. \qquad \text{(eqn 3.11)}$$

The backward shift operator, **B**, is often utilized to express the relationship of the present term to the relevant past terms. The operator is a symbolic indicator and does not imply multiplication of Z_t by a constant **B**. Its use indicates a desire to express past consecutive terms of a series as an ordered group. The operator possesses the following relationship:

$$BZ_t = Z_{t-1}$$
$$B^2 Z_t = Z_{t-2}$$
$$B^m Z_t = Z_{t-m}.$$

Using the operator, the autoregressive process of order p is expressed as

$$\varphi(B)Z_t^* = A_t \qquad \text{(eqn 3.13)}$$

with

$$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p).$$

The moving average model assumes that the series can be expressed as a weighted average of past successive white noise terms. A first order moving average model expressing the series as deviation from the mean of the series is:

$$Z_t^* = A_t - \theta_1 A_{t-1} \qquad \text{(eqn 3.14)}$$

where

$A_t =$ series is white noise $\sim N(0, \sigma_A^2)$

$\theta_1 =$ coefficient of the most recent white noise term.

18

This basic relationship can be expanded to include any number of past terms. A moving average process of order q is written as:

$$Z_t^* = A_t - \theta_1 A_{t-1} - \theta_2 A_{t-2} - ... - \theta_q A_{t-q}. \qquad \text{(eqn 3.15)}$$

Using the backward shift operator this model is expressed as

$$Z_t^* = \theta(B)A_t \qquad \text{(eqn 3.16)}$$

with

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q.$$

A combination of these types of models can be developed and possesses relationships to past series terms and to past white noise elements. The process is called a mixed autoregressive-moving average model and is expressed as:

$$Z_t^* = \varphi_1 Z_{t-1}^* + ... + \varphi_p Z_{t-p}^* + A_t - \theta_1 A_{t-1} - ... - \theta_q A_{t-q}. \qquad \text{(eqn 3.17)}$$

This expression is usually written in terms of the backward shift operator,

$$\varphi(B)Z_t^* = \theta(B)A_t. \qquad \text{(eqn 3.18)}$$

Models which adequately describe the data rarely exhibit values of p or q greater than two and usually are less than two. [Ref. 10: p. 66]

The use of differencing to achieve a stationary series permits the use of the Box-Jenkins method to model series which are nonstationary in nature. The backward differencing operator ∇ is used to indicate the following relationship:

$$\nabla Z_t = Z_t - Z_{t-1} = (1-B)Z_t. \qquad \text{(eqn 3.19)}$$

The model now becomes an autoregressive integrated moving average process (ARIMA) of order (p,d,q,) and is written:

$$\varphi(B)\nabla^d Z_t = \theta(B)A_t. \qquad \text{(eqn 3.20)}$$

In this relationship the superscript d indicates the number of times that regular, backward differencing was utilized to achieve a stationary series.

Seasonal differencing was used in several of the models to achieve stationarity. The inclusion of seasonal differencing produces a seasonal ARIMA model described as $(p,d,q) \times (P,D,Q,)_s$. The subscript s indicates the number of periods contained in one season. In the examination of quarterly data, there are four periods in one year and the subscript becomes a four. The general form of the seasonal ARIMA model is a multiplicative ARIMA model. The multiplicative nature of the model indicates that the model contains terms which are products of the regular and seasonal coefficients. Intuitively, this makes sense. In the case of quarterly data, the value in the series five periods previous to the present is included and has a coefficient which is the negative product of the first order term and the seasonal term [Ref. 10: p. 164]. A multiplicative ARIMA model of order $(1,1,0) \times (1,1,0)_4$ is expressed as:

$$Z_t = \varphi Z_{t-1} + \Phi_1 Z_{t-4} - \varphi_1 \Phi_1 Z_{t-5} + A_t \qquad \text{(eqn 3.21)}$$

with

$$A_t \sim N(0, \sigma_A^2) .$$

Φ_1 = coefficient for the seasonal term.

The general expression for the multiplicative ARIMA model is

$$\varphi_p(B) \Phi_P(B_s) \nabla^d \nabla_s^D Z_t = \theta_q(B) \Theta_Q(B_s) A_t$$

$$\text{(eqn 3.22)}$$

where

Θ_Q = coefficient of the seasonal error term

Φ_P = coefficient for the seasonal term.

Several plausible models are developed during the estimation stage and these alternative models can be compared utilizing the diagnostics provided in most computer packages. The parameters and the associated residuals from each plausible model were examined to validate the model. The residuals were examined for the presence of bias and autocorrelation. The three indicators used to detect the presence of these conditions were the residual mean and variance, the autocorrelations of the residuals, and the Q statistic. The model parameters were examined for statistical

20

significance and indications of high correlation between each other. Highly correlated parameters are usually an indication of the inclusion of unnecessary parameters in the model [Ref. 10: p. 98]. The desired outcome is to determine one or more models which produces fitted values as close as possible to the original series. Additionally, it is desired that the models have as few parameters as possible [Ref. 9: p. 17]. The final stage of the Box-Jenkins method, forecasting, enables the user to project the series into the future. Often. 95% confidence intervals for the projected values are provided by the computer package. The computer package utilized during this analysis was GRAFSTAT. It is a package being developed by IBM and is installed at the Naval Postgraduate School for evaluation.

In general, the results derived from the Box-Jenkins method are usually more accurate in the short and intermediate term than forecasts from other methods, including regression [Ref. 11: p. 236]. The cost of these generally superior results are measured in the computer resources required to derive the model and the expertise required of the statistician to determine an appropriate model. The procedure allows for interpretation on the part of the forecaster. Two forecasters may identify different models as being the best model to fit the same data set. Even so, both sets of forecasts may be highly accurate when compared to the future observations of the series [Ref. 3: p. 11].

The Box-Jenkins models developed for each of the contractors displayed strong predictive properties when used to predict the missing values. Table 3 presents the differencing required to achieve stationarity and the final form of the model for each of the contractors. The model for each contractor is expressed in standard Box-Jenkins notation with a seasonal period of four. The coefficient values, the model mean, and the standard errors of these terms are included in the presentation.

C. BOX-JENKINS TRANSFER FUNCTION

The Box-Jenkins transfer function is a procedure which allows a forecaster to aggregate the information contained in a particular series (output) with one or more related series (input) to forecast future values of the series. The relationship which is usually identified is that the trend present in the input series is reflected in the output series after a lag of several periods. Relationships of this order are referred to as dynamic responses. The aggregation of information is achieved through a transfer

21

TABLE 3
SUMMARY OF BOX-JENKINS MODELS

Summary For Contractor A

ARIMA	$(1,0,0) \times (0,1,0,)_4$
AR(1)	.753
Standard Error	.157
Mean	2023.270
Standard Error	6648.622

Summary For Contractor B

ARIMA	$(1,1,0) \times (0,1,0,)_4$
AR(1)	-.845
Standard Error	.171
Mean	488.515
Standard Error	1162.532

Summary For Contractor C

ARIMA	$(1,2,0) \times (1,0,0)_4$
AR(1)	-.673
Standard Error	.185
SAR(1)	.764
Standard Error	.103
Mean	305.319
Standard Error	416.400

Summary For Contractor D

ARIMA	$(1,0,0) \times (0,0,0,)_4$
AR(1)	.688
Standard Error	.163
Mean	189328.647
Standard Error	6599.820

22

TABLE 3
SUMMARY OF BOX-JENKINS MODELS (CONT'D.)

Summary For Contractor E

ARIMA	$(0,1,1) \times (0,0,0)_4$
MA(1)	.906
Standard Error	.059
Mean	1532.200
Standard Error	397.637

Summary For Contractor F

ARIMA	$(1,1,0) \times (1,1,0,)_4$
AR(1)	-.607
Standard Error	.166
SAR(1)	.875
Standard Error	.129
Mean	-350.716
Standard Error	2534.731

Summary For Contractor G

ARIMA	$(1,0,0) \times (0,1,0)_4$
AR(1)	.854
Standard Error	.103
Mean	1125.453
Standard Error	18744.529

function model [Ref. 9: p. 13]. This procedure is comprised of the same three phases as the univariate Box-Jenkins method. During the identification stage, differencing of the series is usually recommended in an effort to achieve a stationary series. However, a linear combination of elements in the output series often may be stationary, and differencing of all of the series in the model can cause complications in identifying the appropriate model [Ref. 12].

The data trends may be tentatively identified and prewhitened in an effort to achieve an input series which strongly resembles white noise. This effort is made in an

attempt to improve the interpretability of the cross correlation function (CCF). If the input series is not prewhitened, the CCF often cannot be interpreted [Ref. 13: p. 243]. A series which can be represented by an ARIMA model

$$X_t = (1-\varphi_1 B - ... - \varphi_p B^p)^{-1}(1-\theta_1 B - ... - \theta_q B^q)A_t. \qquad \text{(eqn 3.23)}$$

or as expressed with the backward shift operator

$$X_t = \varphi_X^{-1}(B)\theta_X(B)A_t. \qquad \text{(eqn 3.24)}$$

can be prewhitened by inverting the model. The prewhitened version of the series is expressed as follows:

$$\alpha_t = \varphi_X(B)\theta_X^{-1}(B)X_t. \qquad \text{(eqn 3.25)}$$

The same transformation is applied to the output series culminating in the following expression:

$$B_t = \varphi_X(B)\theta_X^{-1}(B)Z_t. \qquad \text{(eqn 3.26)}$$

The relationship between the input and output series is described in an impulse response function of the form

$$Z_t = \upsilon_0 X_t + \upsilon_1 X_{t-1} + \upsilon_2 X_{t-2} + ... + \eta_t \qquad \text{(eqn 3.27)}$$

where

υ is the impulse response

η_t is the random noise term.

Expressing the impulse response function in terms of the prewhitened series, α_t and B_t, the function becomes

$$B_t = \upsilon(B)\alpha_t + \varepsilon_t. \qquad \text{(eqn 3.28)}$$

where ε_t is the transformed noise series defined by

24

$$\epsilon_t = \varphi_X(B)\theta_X^{-1}(B)\eta_t .$$

(eqn 3.29)

The impulse response weights can be determined through examination of the CCF. The following relationship is utilized in this regard:

$$\upsilon_k = \frac{\rho_{\alpha B}(k)S_B}{S_\alpha} \qquad k = 0.1.2.... \quad \text{(eqn 3.30)}$$

where

$\rho_{\alpha B}(k)$ = correlation coefficient between the α_t and B_t series for the kth element

S_α = estimated standard deviation of the α_t series

S_B = estimated standard deviation of the B_t series.

Once estimates of υ_k are obtained, identification of those elements of the impulse function which are statistically significant enables the forecaster to identify an appropriate model to be used as a basis for the transfer function. From these estimates. simultaneous equations can be formed and solved to provide preliminary estimates of the parameters of the model. Since the impulse responses provided by the CCF are statistically inefficient in general, the proposed model is used as a starting point to be fitted by some more elaborate means.

This procedure is extremely complicated and places a high demand on computer resources and the skills of the forecaster. The results should be indicative of the cost of attaining them. In this particular application of the procedure, several difficulties were encountered. A computer package to perform the entire procedure was not available. The GRAFSTAT package mentioned previously does contain a CCF routine. Therefore the inefficient estimates of the impulse responses obtained from the examination of the CCF could be used to form simultaneous equations. However, this would be the most elaborate means of determining the parameters. The data from each contractor were examined in all combinations of undifferenced. regularly differenced, and seasonally differenced forms. In all cases, the data were prewhitened before examination. Only one CCF indicated the presence of an impulse response value which was statistically different from zero. This was the prewhitened and undifferenced series for Contractor D. Therefore. the only input value used in the

25

forecast of the output was the present input value. The nature of this model led to a single equation of the following form:

$$v_0 = 14.39 = \omega_0. \qquad \text{(eqn 3.31)}$$

From equation 3.31, the final multivariate Box-Jenkins model was determined to be:

$$Y_t = 14.39 \, X_t. \qquad \text{(eqn 3.32)}$$

The model parameters are listed in Table 4. The forecasts made with this model are inferior to those obtained from the univariate Box-Jenkins model. As will be discussed in the results chapter, the presence of the input series (X_t) does not produce a regression model or a transfer function which outperforms the univariate Box-Jenkins model for contractor D. The inability of the procedure to develop adequate models for the majority of the contractors may be attributable to the relatively small sample size of each data set. The longest series available for analysis was twenty-seven observations and the majority of the contractors supplied data in the range of twenty to twenty-four observations. Generally, sample sizes larger than sixty observations or data spanning at least eight complete seasonal periods are recommended as a minimum number of data points for the method to perform well [Ref. 3: p. 6].

TABLE 4

SUMMARY FOR CONTRACTOR D

$\rho_{\alpha\beta}$	.49
S_α	1034.734
S_B	30387.40016
v_0	14.39

26

IV. PREDICTION METHODOLOGY AND RESULTS

Each model was used to predict the four withheld data points. For the GLS models actual X's for the four periods were available and used. In those cases where only one form of autocorrelation was present in the data, a transformation of the following type was made to the original data prior to estimation:

$$Y_t^* = \hat{\rho}_i Y_{t-i} + (X_t - \hat{\rho}_i Y_{t-i}) \qquad \text{(eqn 4.1)}$$
$$(t = T-3, T-2, T-1, T)$$
$$(i = 1 \text{ or } 4) .$$

The majority of the GLS model required two transformations to remove all forms of autocorrelation. In those cases, the X's and Y's were transformed in the following manner before the final GLS model was used to predict the withheld values.

Transform the data for AR(4):

$$Y_t^* = Y_t - \hat{\rho}_4 Y_{t-4}$$
$$X_t^* = X_t - \hat{\rho}_4 X_{t-4} \qquad \text{(eqn 4.2)}$$
$$(t = 5, 6, 7, ..., T) .$$

Transform these new data to remove the effects of AR(1):

$$Y_t^{**} = Y_t^* - \hat{\rho}_1 Y_{t-1}^*$$
$$X_t^{**} = X_t^* - \hat{\rho}_1 X_{t-1}^* \qquad \text{(eqn 4.3)}$$
$$(t = 1, 2, 3, ..., T) .$$

Substitution of equation 4.2 into equation 4.3 allows both equations to be combined as follows:

$$Y_t^{**} = Y_t - \hat{\rho}_4 Y_{t-4} - \hat{\rho}_1 (Y_{t-1} - \hat{\rho}_4 Y_{t-5})$$
$$X_t^{**} = X_t - \hat{\rho}_4 X_{t-4} - \hat{\rho}_1 (X_{t-1} - \hat{\rho}_4 X_{t-5}) \qquad \text{(eqn 4.4)}$$
$$(t = 6, 7, 8, ..., T) .$$

27

These equations can be simplified to the following form:

$$Y_t^{**} = Y_t - \widehat{\rho}_1 Y_{t-1} - \widehat{\rho}_4 Y_{t-4} + \widehat{\rho}_1 \widehat{\rho}_4 Y_{t-5}$$
$$X_t^{**} = X_t - \widehat{\rho}_1 X_{t-1} - \widehat{\rho}_4 X_{t-4} + \widehat{\rho}_1 \widehat{\rho}_4 X_{t-5} \qquad \text{(eqn 4.5)}$$
$$(t = 6,7,8,...,T).$$

Assuming that all forms of autocorrelation have been removed, the following relationship holds:

$$Y_t^{**} = X_t^{**} B + \upsilon_t \qquad \text{(eqn 4.6)}$$
$$\upsilon_t \sim N(0,\sigma^2)$$
$$(t = 6,7,8,...,T).$$

The entire transformation can be made in one step and forecasts can be made with the following equation.

$$Y_t = \widehat{\rho}_1 Y_{t-1} + \widehat{\rho}_4 Y_{t-4} - \widehat{\rho}_1 \widehat{\rho}_4 Y_{t-5} +$$
$$(X_t - \widehat{\rho}_1 X_{t-1} - \widehat{\rho}_4 X_{t-4} - \widehat{\rho}_1 \widehat{\rho}_4 X_{t-5}) B \qquad \text{(eqn 4.7)}$$
$$(t = 6,7,8,...,T).$$

This set of equations pertained to the instance in which AR(4) was removed first and AR(1) was removed during the second application of GLS. These equations were developed by substituting the transformation for AR(4) into the equation for the removal of AR(1) processes. A similar expression was derived for those cases in which the correction for AR(1) preceded the removal of AR(4) processes. In the case of the Y's, this procedure required an iterative process to determine the last three values in the Y vector.

The Box-Jenkins models developed for each contractor were used to produce forecasts which were compared to the withheld data points. These forecast values were provided by the GRAFSTAT package. The package also provided ninety-five percent prediction intervals for each forecast. A graphical presentation of the forecast is provided with an analysis of each contractor's data in the remainder of this chapter.

Forecasts were made with the multivariate Box-Jenkins model. These were made by substituting the known values in the input (Xt) series into equation 3.32 and completing the computations.

The forecasting capabilities of each model were measured by three comparison indicators: correlation coefficient between the actual and predicted values, root mean square divided by the mean of the actuals, and the mean absolute percentage error. An analysis of the data and a presentation of the prediction results for each contractor follows.

A.   CONTRACTOR A

Contractor A supplied twenty-two data points. The data was categorized and deflated to constant dollars as specified in Chapter II. Graphical presentations of the raw data is presented in the top two graphs of Figure 4.1 . The upper left hand graph is a presentation of the overhead cost across time (twenty-two consecutive quarters). Likewise, the upper right hand graph is a display of the direct personnel trend across time. The overhead cost versus time graph displays a sharp decline in the first four quarters and a somewhat cyclic increase thereafter. The trend is increasing in general. The direct personnel versus time graph indicates a similar trend in general, but does not appear to be influenced by a seasonal trend to the extent that the overhead cost trend is. The lower left hand graph displays the relationship of overhead cost versus direct personnel. The weak relationship is depicted graphically and in the results of the OLS regression (adjusted R^2 = .183). The lower right hand graph in Figure 4.1 shows the relationship of overhead cost to direct personnel after both series have been transformed to remove the effects of autocorrelation. The adjusted R^2 for this model is .8787 as indicated in Table 3. The graphical portrayal of the transformed data suggests that the GLS model should be a dramatic improvement over the OLS model as a prediction tool. A summary of the predictive results of the GLS model is presented in Table 5. Actual X's were available and used in making predictions for the four withheld quarters.

The Box-Jenkins model developed for Contractor A was presented in Chapter 3. A topic of concern was the amount of data available which could be used as a basis for the model. Generally, the amount of data needed to develop accurate models is fifty observations with one hundred preferred. Models can be developed in the absence of these amounts of data, but the forecaster must utilize experience and past information to develop preliminary models which can be updated as more information becomes available [Ref. 9: p. 18]. A graphical portrayal of the results of the Box-Jenkins model is presented in Figure 4.2. The left hand graph is the actual overhead series including
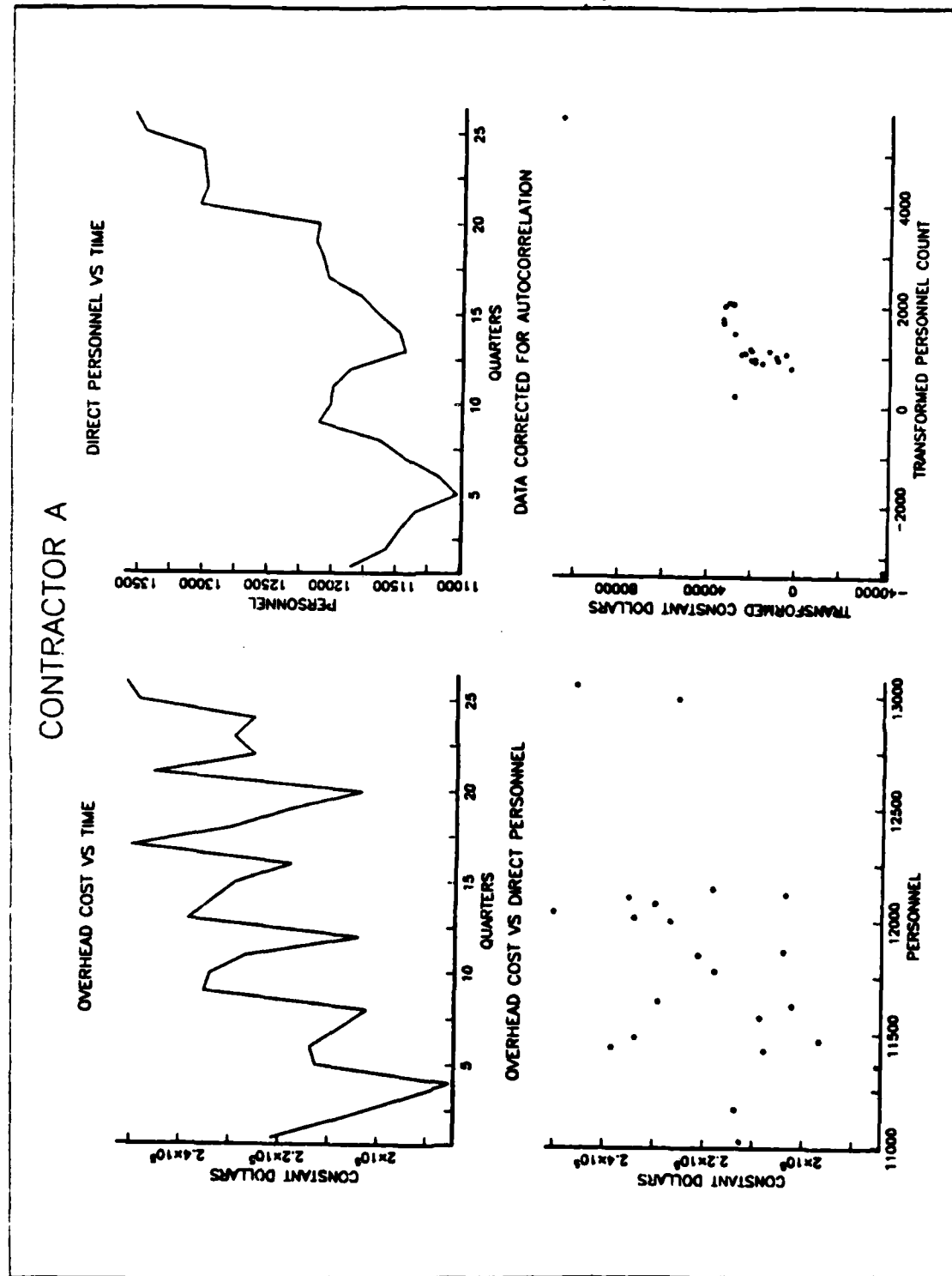
29

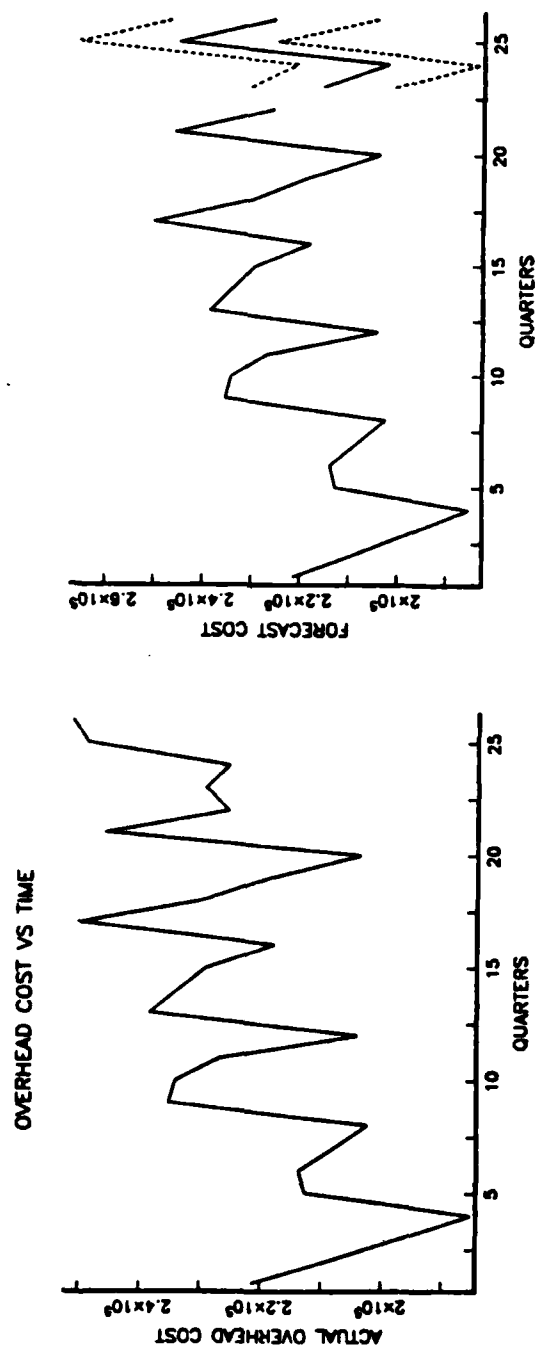Figure 4.1   Regression Analysis Graphs for Contractor A.

Figure 4.2   Box-Jenkins Graphs for Contractor A.

TABLE 5
PREDICTION RESULTS FOR CONTRACTOR A

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	.862	.881
Root Mean Squared Error Mean of the Actuals:	.0649	.0376
Mean Absolute Percentage Error:	5.98	2.79

the periods for which predictions were made. The right hand graph is the first twenty-two observations and the four prediction values. The trend of the predicted values can be compared to the actual trend by mentally superimposing one graph on the other. As indicated in Table 5, the results of the Box-Jenkins model for Contractor A is superior to the results of the GLS model.

Attempts to develop a multivariate Box-Jenkins model were made and proved to be unsuccessful. The cross autocorrelation function (CCF) was plotted. However, none of the impulse weights were statistically significant. All reasonable combinations of differencing were examined in addition to a CCF in which no differencing was included. The outcome was the same in all cases.

B. CONTRACTOR B

Figures 4.3 and 4.4 and Table 6 are provided for Contractor B. Figure 4.3 displays the raw overhead cost and direct personnel series for Contractor B. As indicated in the upper right hand graph, the overhead series has an increasing trend accompanied by a seasonal variation. The direct personnel series is characterized as a consistently increasing series. A comparison of the slopes of the series in each graph indicates that the quarterly direct personnel count is increasing at a slightly greater rate than the overhead cost per period is. The overhead cost versus direct personnel chart reveals the presence of a relationship between those two series which is stronger than the near randomness revealed in the same graph for Contractor A. This is supported
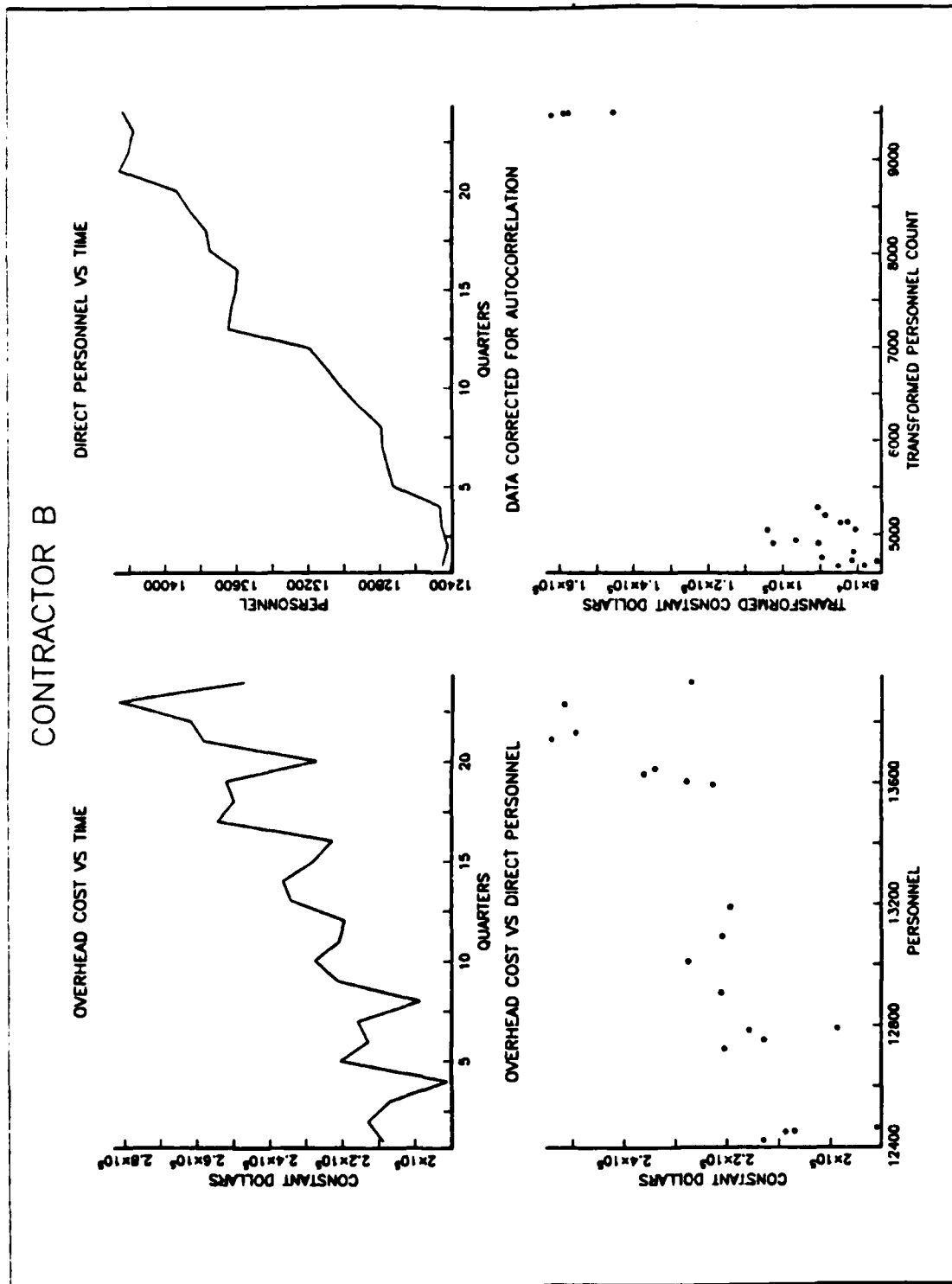
32

Figure 4.3 Regression Analysis Graphs for Contractor B.

Figure 4.4   Box-Jenkins Graphs for Contractor B.
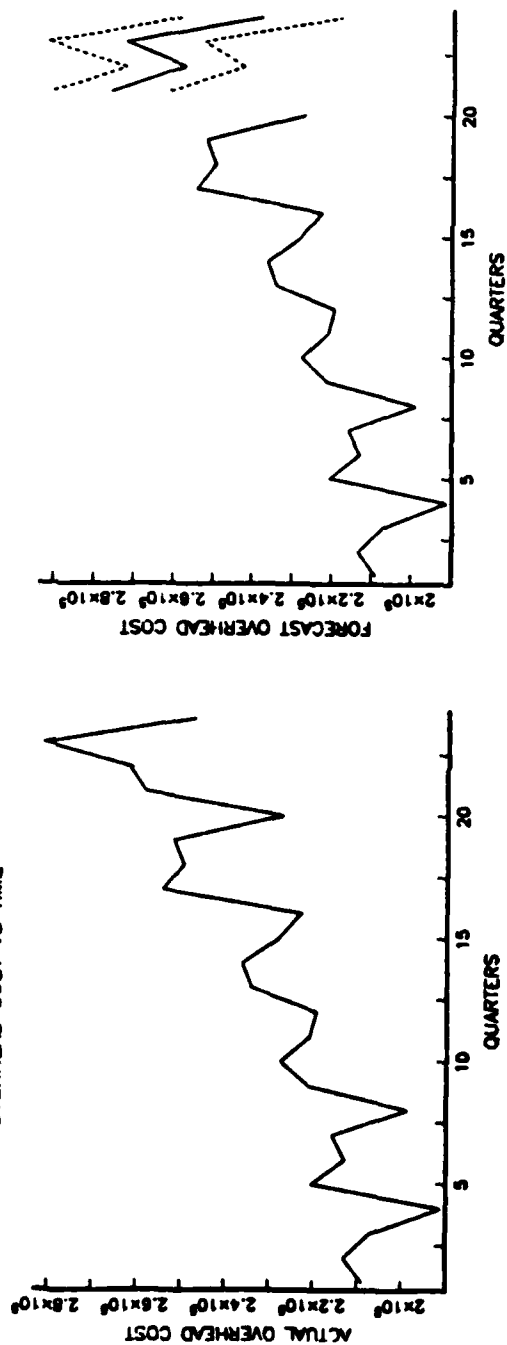
34

TABLE 6

PREDICTION RESULTS FOR CONTRACTOR B

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	.584	.675
Root Mean Squared Error: Mean of the Actuals:	.0952	.0433
Mean Absolute Percentage Error:	8.58	3.91

by the results of the OLS model which are surprisingly good with an adjusted R^2 of .669. After transformation, the GLS model achieves an adjusted R^2 of .927. There are three transformed observations which are easily distinguished from the remainder of the transformed data. These observations are located closely together on the transformed data graph. The GLS model developed possesses all of the indications of a statistically significant and worthwhile prediction model. The adjusted R^2, F-statistic, and T-statistic for the slope are all significant [Ref. 14: p. 133]. Despite this fact, the Box-Jenkins model developed for Contractor B is superior as a forecasting tool in the range which is being examined in this paper. The multivariate Box-Jenkins model suffered from the same short comings as the model for Contractor A. The impulse weights were determined to be statistically insignificant.

C. CONTRACTOR C

Figures 4.5 and 4.6 and Table 7 are provided for Contractor C. The overhead cost and direct personnel graphs derived from Contractor C data are presented in Figure 4.5. The overhead cost trend appears to fluctuate significantly about a mean value of approximately $90,000,000 during the first seventeen quarters with the minimum value of the series occurring in the fifteenth quarter. The series shows a departure from this trend during the last eight quarters. The last four values depicted on the top two graphs are values which were not included during the model development stage. These are the values which are being withheld for prediction
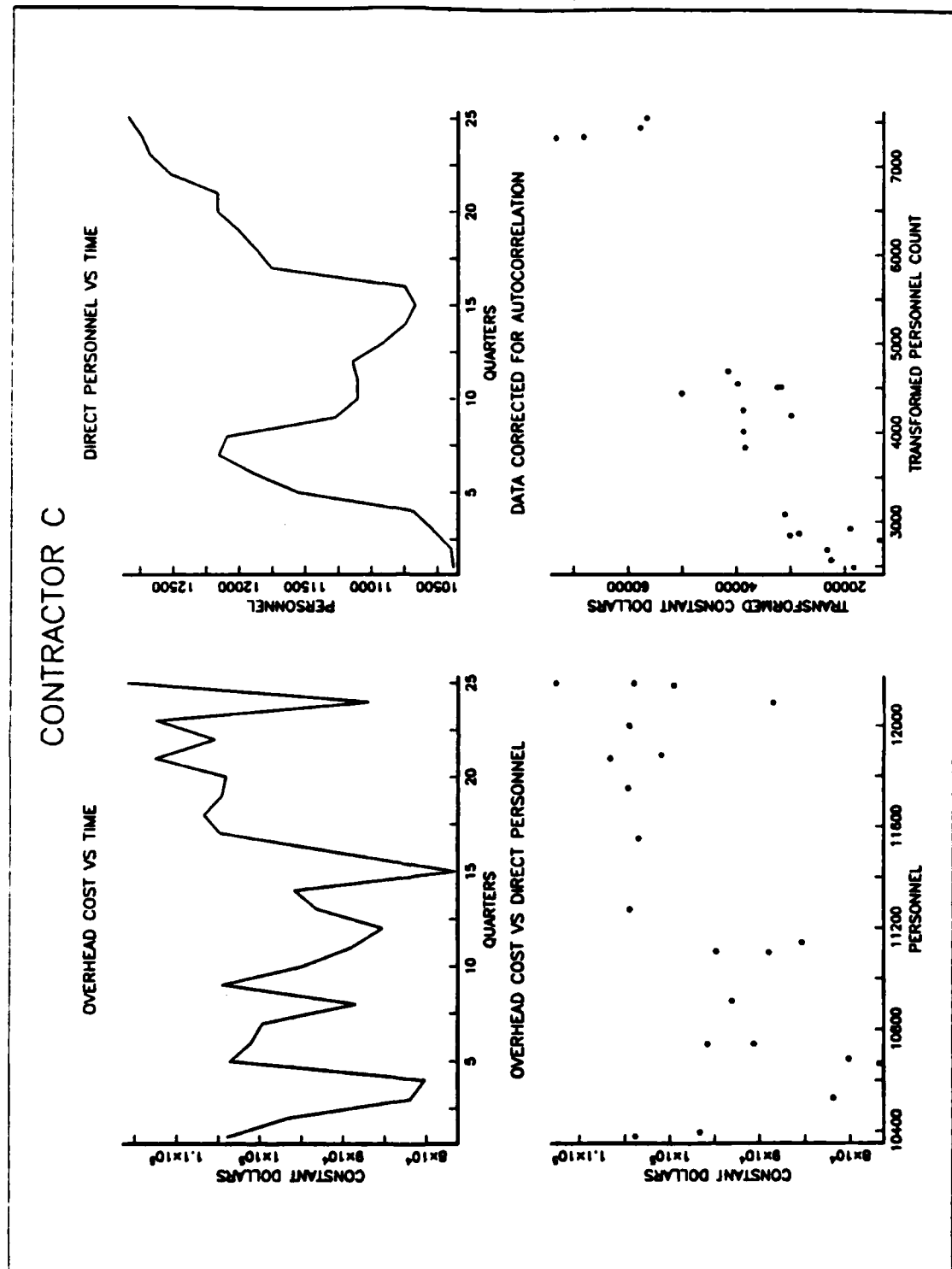
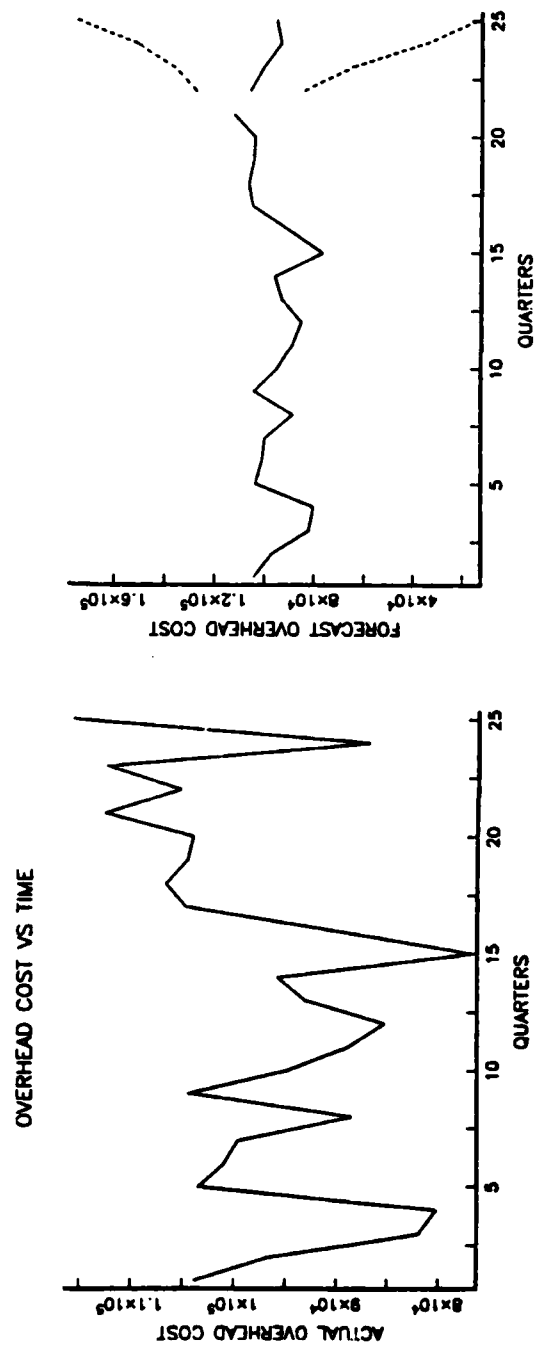Figure 4.5   Regression Analysis Graphs for Contractor C.

Figure 4.6   Box-Jenkins Graphs for Contractor C.

TABLE 7

PREDICTION RESULTS FOR CONTRACTOR C

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	.671	.338
Root Mean Squared Error/ Mean of the Actuals:	.112	.117
Mean Absolute Percentage Error:	8.52	9.10

comparison purposes. The direct personnel graph shows an increase during the first nine periods followed by a decrease for approximately eight periods. At that point, the number of direct workers employed at Contractor C increases for the remaining eight periods. This trend matches the overhead cost trend in general for the entire length of the data strings, but does not appear to be influenced by a seasonal component. The plot of overhead cost versus direct personnel appears nearly random (adjusted R^2 = .291). The plot of the transformed data displays a strong direct relationship between the two variables. As the number of personnel increases, the overhead cost increases. This is apparent in the GLS model which has an adjusted R^2 of .840 and an F statistic of 106.326. The Box-Jenkins model graphs are presented in Figure 4.6 . These graphs are difficult to mentally superimpose on each other. This problem is caused by the bounds of the 95% confidence interval for the forecast observations. In the case of Contractor C, the regression model produces predictions which are actually closer to the actual values than those calculated by the Box-Jenkins model. A comparison of the predictive results is presented in Table 7 . The multivariate Box-Jenkins model was unusable for Contractor C.

D.   CONTRACTOR D

The data supplied by Contractor D spanned twenty periods and is presented in Figures 4.7 and 4.8 and Table 8. The overhead cost versus time plot indicates that the
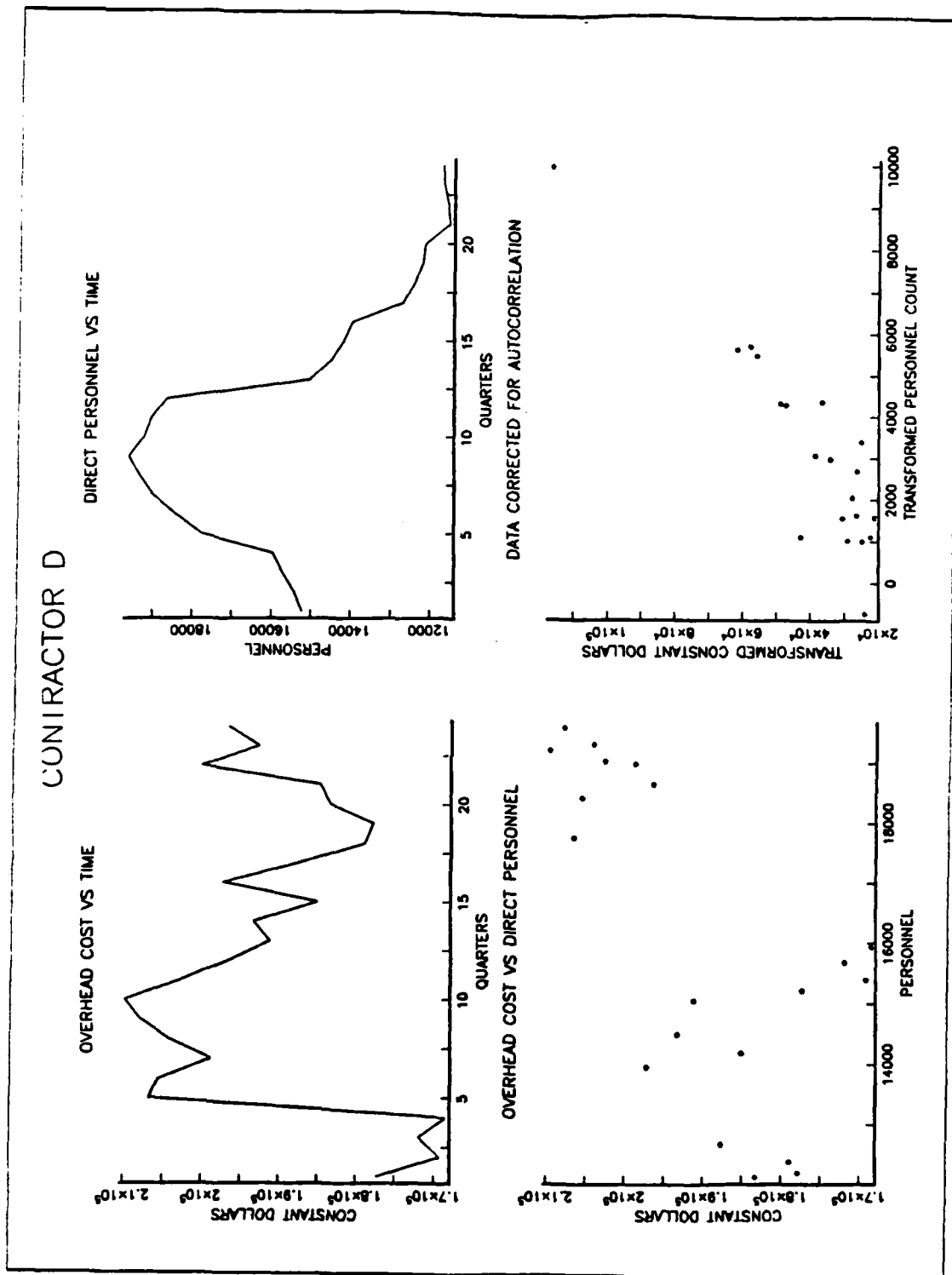
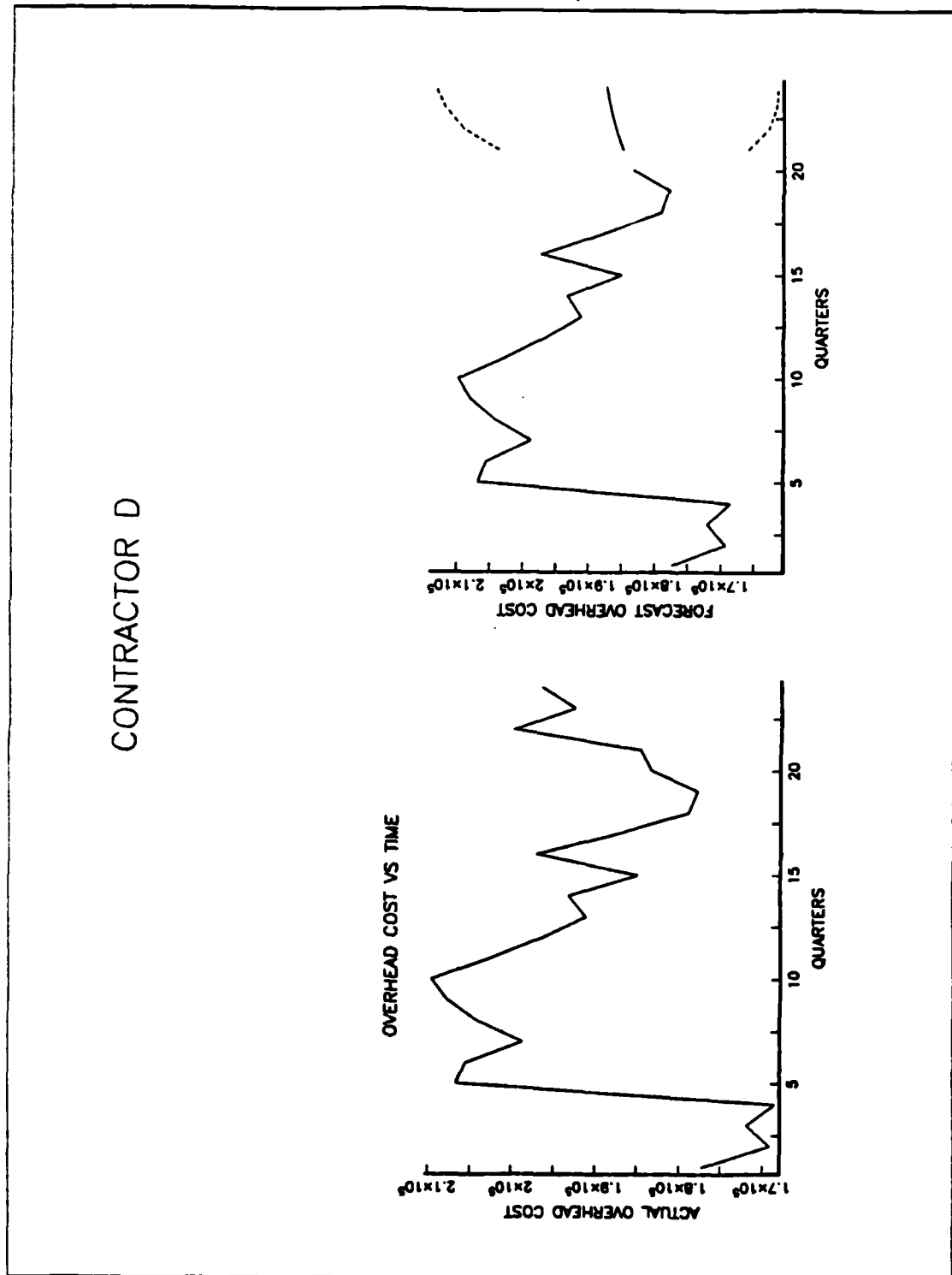Figure 4.7   Regression Analysis Graphs for Contractor D.

Figure 4.8   Box-Jenkins Graphs for Contractor D.

TABLE 8
PREDICTION RESULTS FOR CONTRACTOR D

	REGRESSION	BOX-JENKINS	TRANSFER FUNCTION
Correlation Coefficient:	-.757	.641	.418
Root Mean Squared Error Mean of the Actuals:	.197	.0435	.126
Mean Absolute Percentage Error:	18.64	3.54	12.18

series started at a relatively low point and increased rapidly in the fifth quarter. The trend remained relatively constant for six periods at which time it began to decrease slowly. This trend remained consistent through the twentieth quarter which is the last period included in the model development stage. The direct personnel series is similar to the overhead cost trend with an abrupt decrease in the thirteenth quarter. Beginning in the twenty-first quarter, the direct personnel trend becomes flat and neither increases nor decreases for the remainder of the series. These last four values are the known X's which are used in the regression model and the multivariate Box-Jenkins models. As the X's (direct personnel) become a level function in the prediction interval of the data, the Y's (overhead cost) suddenly increases. This departure from the previous trend causes problems with both of the models which utilize the available X values to generate predictions. The effect of this departure from the trend is indicated by the relatively poor results of these models as listed in Table 8. The multivariate Box-Jenkins model is predicated on a single impulse weight which was significant. This occurred at a lag of zero periods. The model which was developed utilized only the present period X value to predict the value of Y. Therefore, this model was susceptible to the trend departures present in the data.

E. CONTRACTOR E

Figures 4.9 and 4.10 and Table 9 are provided for Contractor E. The regression model developed for Contractor E is unique in the fact that it did not need to be corrected for autocorrelation. As indicated in the graphs in Figure 4.9, both the
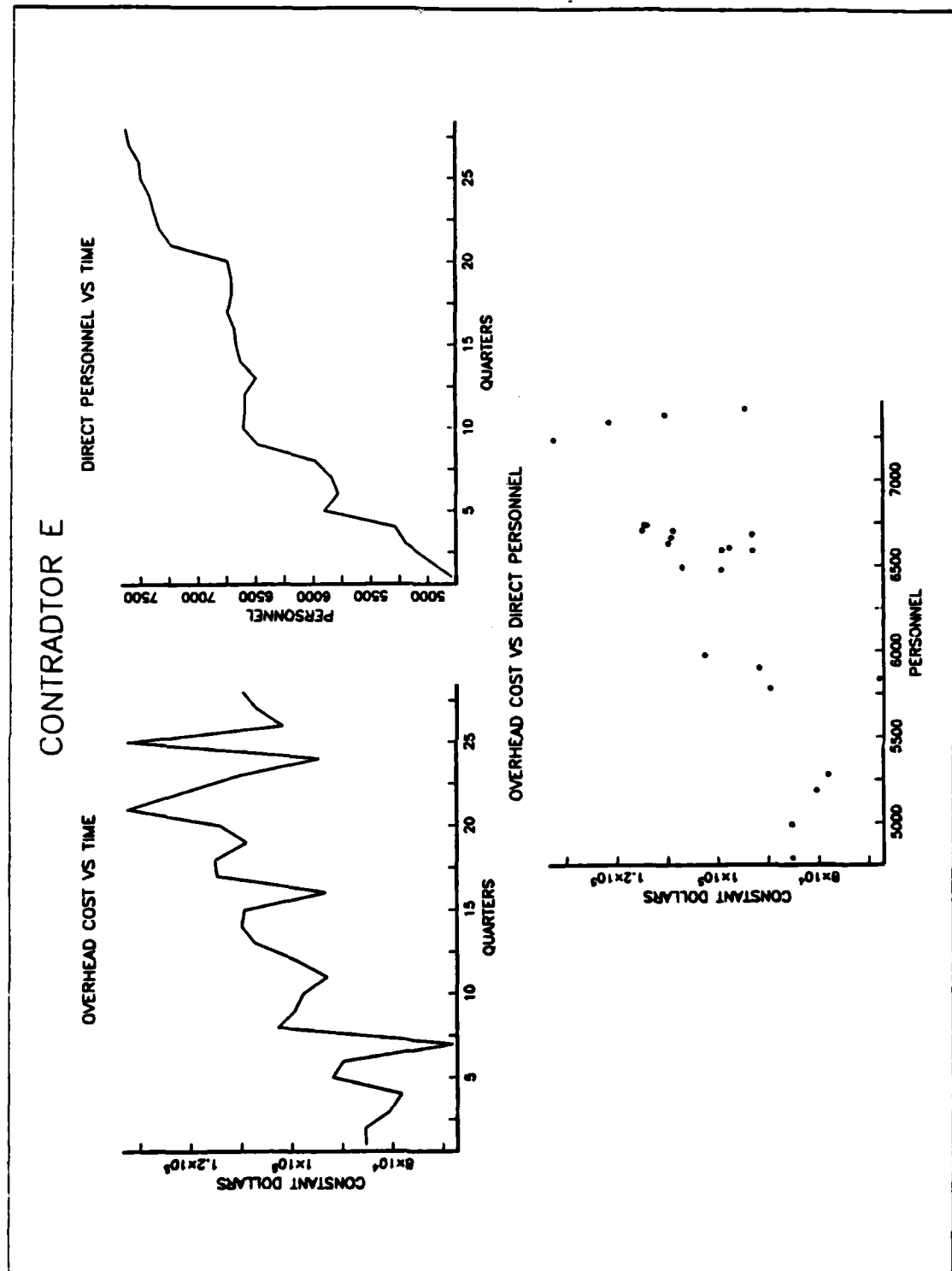
Figure 4.9   Regression Analysis Graphs for Contractor E.

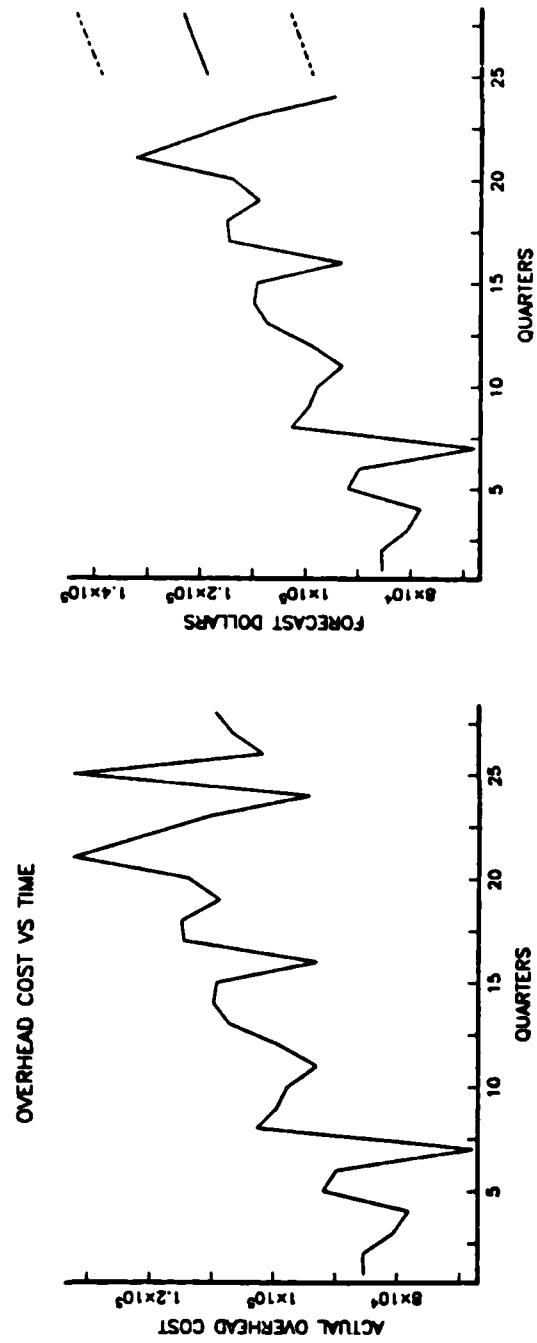Figure 4.10   Box-Jenkins Graphs for Contractor E.

43

TABLE 9

PREDICTION RESULTS FOR CONTRACTOR E

CONTRACTOR E

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	-.434	-.606
Root Mean Squared Error: Mean of the Actuals:	.107	.137
Mean Absolute Percentage Error:	8.99	13.83

overhead cost and the direct personnel series are generally increasing with time. The overhead cost series tends to fluctuate above and below the general trend throughout the length of the series with the most significant departure from the general trend occurring in the twenty-first to the twenty-fifth data observations. The direct personnel series increased initially in the first ten periods and remained fairly constant at a level of approximately 6500 workers for the next ten periods. This series shows a sharp increase in the twenty-first period and continues to increase at a somewhat slower rate thereafter. The overhead cost versus direct personnel graph displays a direct relationship between these two variables in the absence of an autocorrelation correction. The strength of this relationship is less apparent as the number of personnel increases. The four observations in the upper right hand corner of this graph are the last four observations in each series in the model formulation range. Their proximity to each other is a function of the fact that the direct personnel series is slowly increasing in the twenty-first through the twenty-fourth observations. This explains the remoteness of their placement on the graph as they occur later in time than the large increase in the number of workers that was recorded in the twenty-first quarter. The appearance of these points as a nearly vertical line is a function of the small increase in the direct personnel component of the graph and the large fluctuations that occurred in the overhead cost series in the twenty-first to twenty-fourth quarters. The OLS model has an R^2 of .527 and an F statistic of 26.59.

The ARIMA model for Contractor E was also unique in the fact that it was the only contractor model which utilized a moving average (MA) process to describe the data trend. This model was difficult to identify and was actually determined through a process of elimination. Every reasonable model was analyzed and the ARIMA $(0,1,1) \times (0,0,0)_4$ was chosen on the basis of the statistical significance of the coefficient and the model's performance as a forecasting tool. Several models which actually produced slightly better prediction results were excluded from consideration because the resulting coefficients were not significant. The multivariate Box-Jenkins model could not be developed. The prediction results of the regression and Box-Jenkins models are presented in Table 9 .

F. CONTRACTOR F

Contractor F was the only contractor to supply data covering the entire observation period. Figures 4.11 and 4.12 and Table 10 are provided for Contractor F. The overhead cost and direct personnel series display similar trends. Both series increase until the ninth period, then decrease in general and eventually resume increasing. The overhead cost trend resumes increasing in the sixteenth quarter. This latter trend in the overhead cost series displays a seasonal fluctuation with the generally increasing trend. The direct personnel series decreases until the sixteenth period at which point it becomes a nearly constant function for six quarters. Beginning in the twenty-fourth quarter, the number of workers begins a rapidly increasing trend and continues in this manner through the remainder of the series. Several significant single increases and decreases occur within the direct personnel trend. Three of these rapid changes in the number of direct workers employed by Contractor F border above and below the 3600 to 3800 interval in the work force level. The absence of data observations in this range creates two distinct clusters of observations in the overhead cost versus direct personnel graph. This plot does not appear random because of the blank interval between the two groupings. However, the plot does not display a strong relationship between the components either. The OLS model has an R^2 of .317. Both forms of autocorrelation, AR(1) and AR(4), were removed from the raw data through transformation and the resulting series displays a direct relationship as depicted in the lower right hand graph of Figure 4.11 The application of GLS to the raw data improves the model significantly.
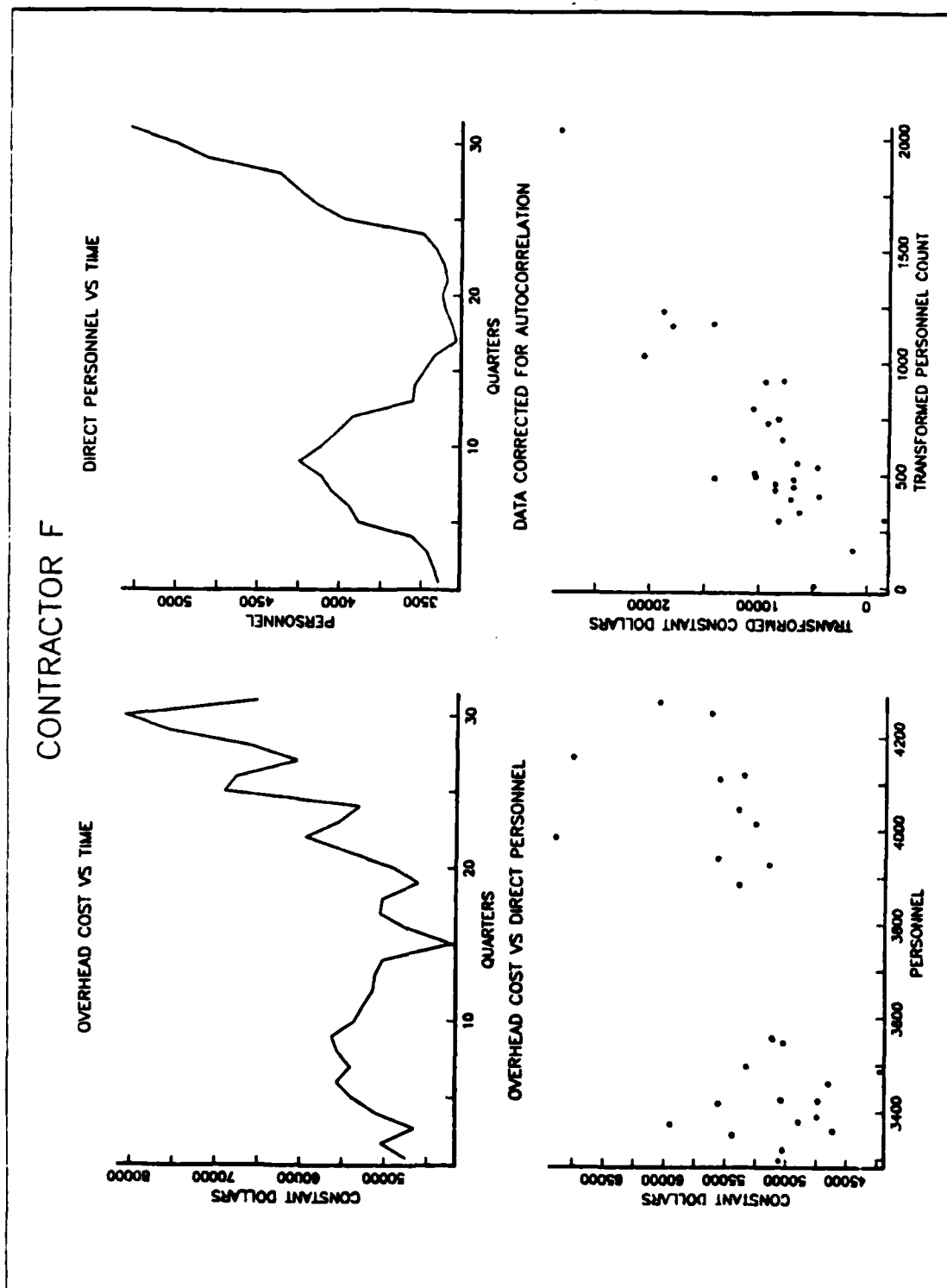
45

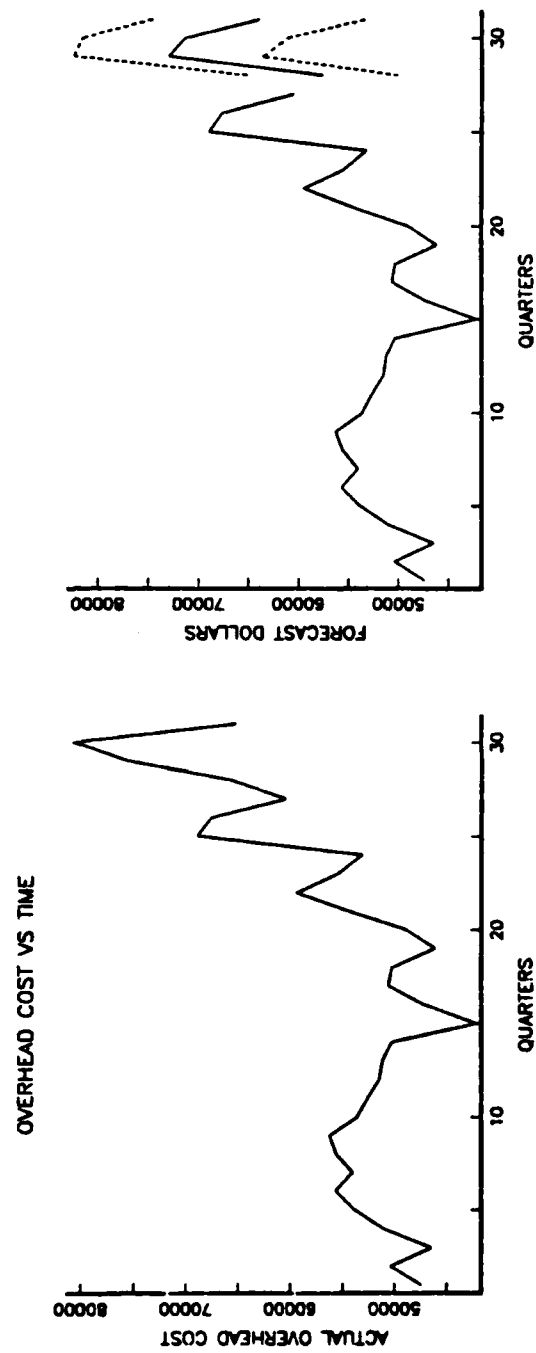Figure 4.11   Regression Analysis Graphs for Contractor F.

Figure 4.12   Box-Jenkins Graphs for Contractor F.

TABLE 10

PREDICTION RESULTS FOR CONTRACTOR F

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	.623	.440
Root Mean Squared Error/ Mean of the Actuals:	.0886	.0823
Mean Absolute Percentage Error:	8.23	6.86

The Box-Jenkins model appears to approximate the trend well but underestimates the magnitude of the actual values. This model is approximately equal to the GLS model in prediction power. The attempted development of the multivariate Box-Jenkins model for Contractor F failed to indicate a significant impulse weight. Therefore, the development of the model was not pursued.

G.   CONTRACTOR G

The data supplied by Contractor G spanned twenty-four quarters. The graphs in Figure 4.13 indicate that both the overhead cost and direct personnel series follow similar patterns during the twenty-four quarters. There is an increasing trend which becomes a decreasing trend in the vicinity of the seventh quarter for both series. The overhead cost series displays the influence of seasonal fluctuations about the general trend. The lower left hand plot of overhead cost versus direct personnel indicates that the uncorrected data possess a noticable direct relationship to each other. The R^2 value for the OLS model was .541. After the data were corrected for both AR(1) and AR(4) processes resulting in a GLS model with an R^2 of .862. The F statistic for this model was 133.155. Table 11 illustrates the outstanding predictive results of this model. The Box-Jenkins model for Contractor G is portrayed graphically in Figure 4.14. The scaling of the two graphs is different because of the inclusion of the ninety-five percent confidence intervals in the forecast graph. As indicated in Table 11, the
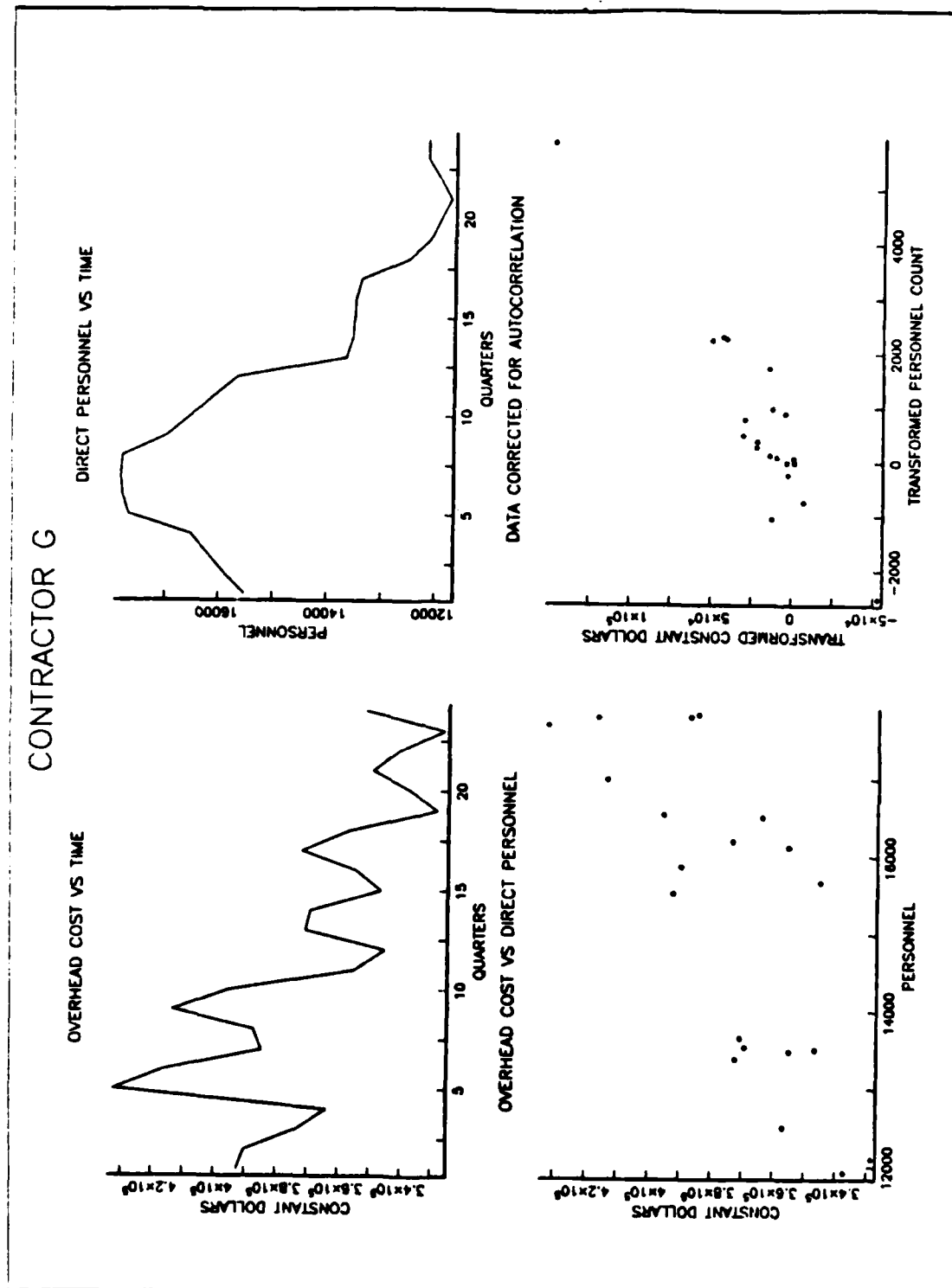
48

Figure 4.13   Regression Analysis Graphs for Contractor G.

49

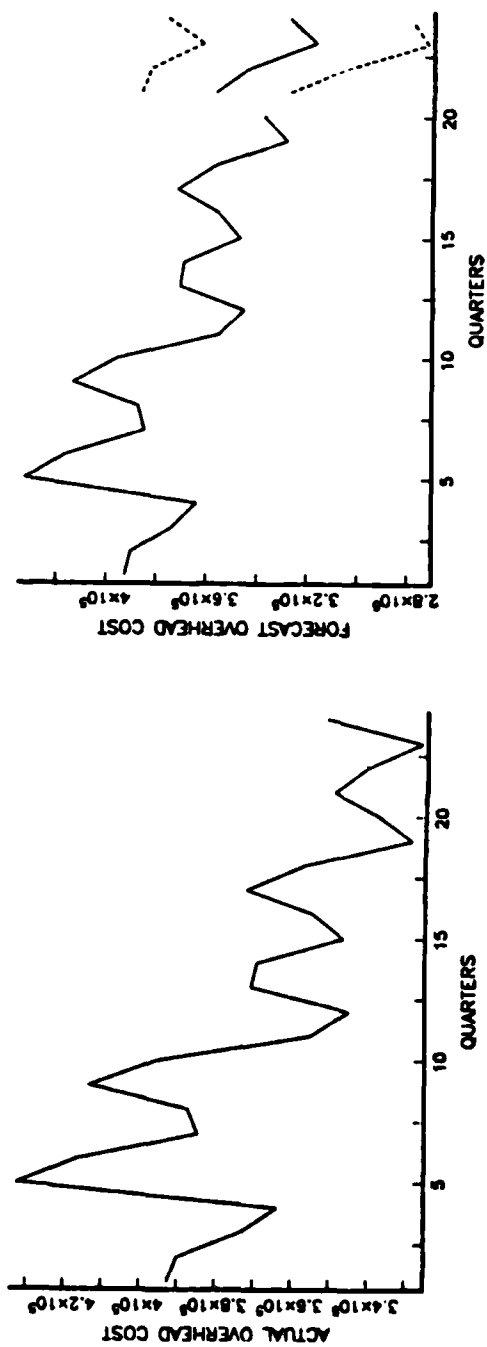Figure 4.14   Box-Jenkins Graphs for Contractor G.

50

TABLE 11

PREDICTION RESULTS FOR CONTRACTOR G

	REGRESSION	BOX-JENKINS
Correlation Coefficient:	.638	.604
Root Mean Squared Error Mean of the Actuals:	.0262	.0394
Mean Absolute Percentage Error:	1.96	3.15

model is an effective forecasting tool for the specified prediction range. However, the regression model is superior in all three categories of comparison. The multivariate Box-Jenkins model could not be developed for Contractor G. The recurring problem of insignificant impulse weights plagued this data set as it has several others examined in this thesis.

V. CONCLUSIONS

The intent of this project was to develop and compare forecasting models to be used in the prediction of overhead costs for seven government aerospace contractors. This is part of a continuing overhead tracking project at the Naval Air Systems Command. Three types of methodology were considered as possible model sources. These included least squares regression models, Box-Jenkins methods, and Box-Jenkins transfer functions.

The regression models which were examined were developed to be used by unsophisticated forecasters and operate well on a microcomputer. The characteristics of the data suggested the complications in model development which are associated with autocorrelation. Consequently, these models test for two forms of autoregressive processes: AR(1) and AR(4). Any necessary adjustments are made within the model. A review of several literature sources indicated that the predictive capabilities of regression models are usually inferior to those of the Box-Jenkins method in the short run. In general this was found to be true. However, the regression models performed well in most cases and resulted in superior forecasting models for two of the contractors.

The Box-Jenkins class of models is considerably more complex than least squares regression. Therefore, a more experienced forecaster and more efficient computer resources are required to employ this method. The majority of the Box-Jenkins transfer function models could not be developed for the data in this project. The difficulties encountered are probably due to the small sample size for each data set. Additionally, a computer package was not available which could perform the entire transfer function procedure.

In those cases in which these requirements could be fulfilled, either of the Box-Jenkins methods should be utilized. In the absence of continuous access to such capabilities such as is the case in the Naval Air Systems Command overhead tracking project, least squares regression theory can be used to develop very adequate models which produce results which are usually far superior to the results obtained from the prevailing practice of applying estimated overhead rates to estimated direct labor hours.

APPENDIX
APL FUNCTIONS

The following APL functions perform the OLS and GLS regressions. The GLS function transforms the data to remove AR(4) processes and GLSD makes the required transformation for AR(1). The PRED function computes the prediction effectiveness measurements for the OLS, Box-Jenkins, and transfer function forecasts. The PRED1, PRED4, and PRED40 functions perform the same task for the regression models which required a transformation due to the presence of autocorrelation. The PRED1 function was utilized in those cases where only AR(1) processes were detected during the regression model development stage. The PRED4 function performs a similar task for those cases in which AR(4) processes are the only form of autocorrelation. The PRED40 function makes the appropriate transformations for both AR(1) and AR(4) processes regardless of the order in which they were removed.

```
     - OLS;X2;Y;N;X1;XT;X;XTXI;XTY;K1;K;YHAT;YBAR;AY;ERR;ET14
[1]  ⍝THIS FUNCTION COMPUTES THE REGRESSION STATISTICS OF AN
[2]  ⍝INDEPENDENT VARIABLE AND A DEPENDENT VARIABLE.
[3]  ⍝THE FUNCTION SUPPLIES THE COLUMN OF ONES.
[4]  'ENTER THE INDEPENDENT VARIABLE'
[5]  DL←⎕
[6]  'ENTER THE DEPENDENT VARIABLE'
[7]  OH←⎕
[8]  X2←DL
[9]  Y←OH
[10] N←⍴Y
[11] Y←(N,1)⍴Y
[12] X1←N⍴1
[13] XT←(2,N)⍴(X1,X2)
[14] X←⍉XT
[15] XTXI←⌹(XT+.×X)
[16] XTY←XT+.×Y
[17] B←XTXI+.×XTY
[18] K1←⍴B
```

53

```
[19] K←K1[1]
[20] YHAT←X+.×B
[21] E←Y-YHAT
[22] RSS←1ρ((⍉E)+.×E)
[23] S2←RSS÷(N-K)
[24] S←S2*0.5
[25] YBAR←(+/((N)ρY))÷N
[26] AY←((N)ρY)-YBAR
[27] TSS←((N)ρY)+.×AY
[28] ESS←TSS-RSS
[29] R2←ESS÷TSS
[30] ADJR2←1-((RSS÷(N-K))÷(TSS÷(N-1)))
[31] COVB←S2×XTXI
[32] STERRA←COVB[1;1]*0.5
[33] STERRB←COVB[2;2]*0.5
[34] TA←B[1;1]÷(COVB[1;1]*0.5)
[35] TB←B[2;1]÷(COVB[2;2]*0.5)
[36] F←(R2÷(K-1))÷((1-R2)÷(N-K))
[37] ERR←(N)ρE
[38] ET1←0,((N-1)↑ERR)
[39] EET1←(1↓(ERR-ET1))*2
[40] D1←1ρ((+/EET1)÷RSS)
[41] ET4←0 0 0 0,((N-4)↑ERR)
[42] EET4←(4↓(ERR-ET4))*2
[43] D4←1ρ((+/EET4)÷RSS)
[44] P41←1-(0.5×D4)
[45] P42←(((N*2)×(1-(0.5×D4)))+(K*2))÷((N*2)-(K*2))
[46] P10←(+/(1↓(ERR×ET1)))÷(+/(1↓⁻1↓(ERR×ERR)))
[47] 'RSS    ',⍕RSS
[48] 'ESS    ',⍕ESS
[49] 'TSS    ',⍕TSS
[50] 'S      ',⍕S
[51] '       '
[52] 'ADJR2    ',⍕ADJR2
[53] '        '
```

54

```
[54] 'A      ',(⍕B[1;])
[55] 'STERRA   ',⍕STERRA
[56] 'B      ',(⍕B[2;])
[57] 'STERRB   ',⍕STERRB
[58] '       '
[59] 'F      ',⍕F
[60] 'TA      ',⍕TA
[61] 'TB      ',⍕TB
[62] '       '
[63] 'D1      ',⍕D1
[64] 'P10      ',⍕P10
[65] '       '
[66] 'D4      ',⍕D4
[67] 'P41      ',⍕P41
[68] 'P42      ',⍕P42
      ∇
```

```
    - GLS;X2;Y;N;Y1;Y2;Y1G;Y2G;X21;X22;X21G;X22G;X1;XT;X
-1- ⍝THIS FUNCTION TRANSFORMS THE DATA TO REMOVE THE EFFECTS
-2- ⍝OF AR(4).  THE VECTORS OF XG AND YG ARE THE
-3- ⍝TRANSFORMED VECTORS AND ARE GLOBAL VARIABLES.
[4]  'ENTER THE INDEPENDENT VARIABLE'
[5]  A←□
[6]  'ENTER THE DEPENDENT VARIABLE'
[7]  Z←□
[8]  'ENTER THE ESTIMATE OF P4'
[9]  P4←□
[10] X2←A
[11] Y←Z
[12] N←ρY
[13] Y1←4↑Y
[14] Y2←(N-4)↑Y
[15] Y1G←Y1×((1-(P4*2))*0.5)
[16] Y2G←(4↓Y)-(P4×Y2)
```

```
[17]  YG←Y1G,Y2G
[18]  Y←(N,1)ρYG
[19]  X21←4↑X2
[20]  X22←(N-4)↑X2
[21]  X21G←X21×((1-(P4*2))*0.5)
[22]  X22G←(4↓X2)-(P4×X22)
[23]  XG←(X21G,X22G)
[24]  X1←Nρ1
[25]  XT←(2,N)ρ(X1,XG)
[26]  X←⍉XT
[27]  XTXI←⌹(XT+.×X)
[28]  XTY←XT+.×Y
[29]  BG←XTXI+.×XTY
[30]  K1←ρBG
[31]  K←K1[1]
[32]  YHATG←X+.×BG
[33]  EG←Y-YHATG
[34]  RSSG←1ρ((⍉EG)+.×EG)
[35]  S2G←RSSG÷(N-K)
[36]  SG←S2G*0.5
[37]  YBARG←(+/((N)ρYG))÷N
[38]  AYG←((N)ρYG)-YBARG
[39]  TSSG←((N)ρYG)+.×AYG
[40]  ESSG←TSSG-RSSG
[41]  R2G←ESSG÷TSSG
[42]  ADJR2G←1-((RSSG÷(N-K))÷(TSSG÷(N-1)))
[43]  COVBG←S2G×XTXI
[44]  STERRAG←COVBG[1;1]*0.5
[45]  STERRBG←COVBG[2;2]*0.5
[46]  TAG←BG[1;1]÷(COVBG[1;1]*0.5)
[47]  TBG←BG[2;1]÷(COVBG[2;2]*0.5)
[48]  FG←(R2G÷(K-1))÷((1-R2G)÷(N-K))
[49]  ERRG←(N)ρEG
[50]  ET1G←0,((N-1)↑ERRG)
[51]  EET1G←(1↓(ERRG-ET1G))*2
```

56

```
[52] D1G←1ρ((+/EET1G)÷RSSG)
[53] ET4G← 0 0 0 0 ,((N-4)↑ERRG)
[54] EET4G←(4↓(ERRG-ET4G))*2
[55] D4G←1ρ((+/EET4G)÷RSSG)
[56] P41G←1-(0.5×D4G)
[57] P42G←(((N*2)×(1-(0.5×D4G)))+(K*2))÷((N*2)-(K*2))
[58] P1G←(+/(1↓(ERRG×ET1G)))÷(+/(1↓¯1↓(ERRG×ERRG)))
[59] 'RSSG    ',⍕RSSG
[60] 'ESSG    ',⍕ESSG
[61] 'TSSG    ',⍕TSSG
[62] 'SG      ',⍕SG
[63] '        '
[64] 'ADJR2G  ',⍕ADJR2G
[65] '        '
[66] 'AG      ',(⍕BG[1;])
[67] 'STERRAG ',⍕STERRAG
[68] 'BG      ',(⍕BG[2;])
[69] 'STERRBG ',⍕STERRBG
[70] '        '
[71] 'FG      ',⍕FG
[72] 'TAG     ',⍕TAG
[73] 'TBG     ',⍕TBG
[74] '        '
[75] 'D1G     ',⍕D1G
[76] 'P1G     ',⍕P1G
[77] '        '
[78] 'D4G     ',⍕D4G
[79] 'P41G    ',⍕P41G
[80] 'P42G    ',⍕P42G
      ∇
```

```
    ∇ GLSD;X2;Y;N;Y1;Y2;X21;X22;X1;XT;X;XTXI;XTY;K1;K
[1] ⍝THIS FUNCTION CORRECTS THE REGRESSION MODEL FOR THE
[2] ⍝THE EFFECTS OF AR(1). THE TRANSFORMED VECTORS XF AND
```

57

```
[3]   ⍝YF ARE GLOBAL VARIABLES.
[4]   'ENTER THE INDEPENDENT VARIABLE'
[5]   C←□
[6]   'ENTER THE DEPENDENT VARIABLE'
[7]   D←□
[8]   'ENTER THE VALUE OF P1'
[9]   P1←□
[10]  X2←C
[11]  Y←D
[12]  N←⍴Y
[13]  Y1←0,1↑Y
[14]  Y2←(N-1)↑Y
[15]  Y1F←Y1×((1-(P1*2))*0.5)
[16]  Y2F←(1↓Y)-(P1×Y2)
[17]  YF←1↓(Y1F,Y2F)
[18]  Y←(N,1)⍴YF
[19]  X21←0,1↑X2
[20]  X22←(N-1)↑X2
[21]  X21F←X21×((1-(P1*2))*0.5)
[22]  X22F←(1↓X2)-(P1×X22)
[23]  XF←1↓(X21F,X22F)
[24]  X1←N⍴1
[25]  XT←(2,N)⍴(X1,XF)
[26]  X←⍉XT
[27]  XTXI←⌹(XT+.×X)
[28]  XTY←XT+.×Y
[29]  BF←XTXI+.×XTY
[30]  K1←⍴BF
[31]  K←K1[1]
[32]  YHATF←X+.×BF
[33]  EF←Y-YHATF
[34]  RSSF←1⍴((⍉EF)+.×EF)
[35]  S2F←RSSF÷(N-K)
[36]  SF←S2F*0.5
[37]  YBARF←(+/((N)⍴YF))÷N
```

58

```
[38] AYF←((N)ρYF)-YBARF
[39] TSSF←((N)ρYF)+.×AYF
[40] ESSF←TSSF-RSSF
[41] R2F←ESSF÷TSSF
[42] ADJR2F←1-((RSSF÷(N-K))÷(TSSF÷(N-1)))
[43] COVBF←S2F×XTXI
[44] STERRAF←COVBF[1;1]*0.5
[45] STERRBF←COVBF[2;2]*0.5
[46] TAF←BF[1;1]÷(COVBF[1;1]*0.5)
[47] TBF←BF[2;1]÷(COVBF[2;2]*0.5)
[48] FF←(R2F÷(K-1))÷((1-R2F)÷(N-K))
[49] ERRF←(N)ρEF
[50] ET1F←0,((N-1)↑ERRF)
[51] EET1F←(1↓(ERRF-ET1F))*2
[52] D1F←1ρ((+/EET1F)÷RSSF)
[53] ET4F← 0 0 0 0 ,((N-4)↑ERRF)
[54] EET4F←(4↓(ERRF-ET4F))*2
[55] D4F←1ρ((+/EET4F)÷RSSF)
[56] P41F←1-(0.5×D4F)
[57] P42F←(((N*2)×(1-(0.5×D4F)))+(K*2))÷((N*2)-(K*2))
[58] P1F←(+/(1↓(ERRF×ET1F)))÷(+/(1↓¯1↓(ERRF×ERRF)))
[59] 'RSSF    ',⍕RSSF
[60] 'ESSF    ',⍕ESSF
[61] 'TSSF    ',⍕TSSF
[62] 'SF     ',⍕SF
[63] '      '
[64] 'ADJR2F   ',⍕ADJR2F
[65] '      '
[66] 'AF     ',(⍕BF[1;])
[67] 'STERRAF  ',⍕STERRAF
[68] 'BF     ',(⍕BF[2;])
[69] 'STERRBF  ',⍕STERRBF
[70] '      '
[71] 'FF     ',⍕FF
[72] 'TAF     ',⍕TAF
```

59

```
[73] 'TBF    ',⍕TBF
[74] '        '
[75] 'D1F    ',⍕D1F
[76] 'P1F    ',⍕P1F
[77] '        '
[78] 'D4F    ',⍕D4F
[79] 'P41F   ',⍕P41F
[80] 'P42F   ',⍕P42F
     ∇


     ∇ PRED;YPA;YAA;YPM;YAM;NUM;DENOM
[1] ⍝THIS FUNCTION PROVIDES THE MEASURES OF PREDICTIVE
[2] ⍝EFFECTIVENESS FOR THE BOX-JENKINS MODELS.
[3] 'ENTER THE ACTUAL SERIES VECTOR'
[4] OH2←⎕
[5] 'ENTER THE BOX-JENKINS FORECAST VALUES'
[6] YP←⎕
[7] YA←¯4↑OH2
[8] N←ρYA
[9] MAPE←((+/(((|(YP-YA))÷YA))÷N)×100
[10] RMSE←(((+/((YP-YA)*2))÷N)*0.5)÷((+/YA)÷N)
[11] YPA←(+/YP)÷N
[12] YAA←(+/YA)÷N
[13] YPM←YP-YPA
[14] YAM←YA-YAA
[15] NUM←(+/(YPM×YAM))÷N
[16] DENOM←(((+/(YPM*2))÷N)×((+/(YAM*2))÷N))*0.5
[17] CORR←NUM÷DENOM
[18] 'PREDICTED Y'
[19] YP
[20] '    '
[21] 'ACTUAL Y'
[22] YA
[23] '    '
```

60

```
[24] 'CORRELATION   ',⍕CORR
[25] '     '
[26] 'RMSE/MA      ',⍕RMSE
[27] '     '
[28] 'MAPE          ',⍕MAPE
      ∇


      ∇ PRED1;YPA;YAA;YPM;YAM;NUM;DENOM
[1] ⍝THIS FUNCTION COMPUTES THE MEASURES OF PREDICTIVE
[2] ⍝EFFECTIVENESS FOR THOSE CASES IN WHICH CORRECTIONS
[3] ⍝FOR FIRST ORDER AUTOCORRELATION ARE MADE.'
[4] 'ENTER THE VECTOR CONTAINING THE ACTUAL Y VALUES.'
[5] OH2←⎕
[6] 'ENTER THE VECTOR CONTAINING THE ACTUAL X VALUES.'
[7] DL2←⎕
[8] 'ENTER THE VALUE OF P1.'
[9] P←⎕
[10] YA←¯4↑OH2
[11] X←¯4↑DL2
[12] XN←4↑(¯5↑DL2)
[13] YN←5⍴0
[14] YP←4⍴0
[15] YN[1]←1↑(¯5↑OH2)
[16] L←0
[17] ITERATE:L←L+1
[18] YP[L]←YN[L+1]←(P×YN[L])+((X[L]-(P×XN[L]))×B)
[19] →(L≤3)/ITERATE
[20] YN←4↑YN
[21] N←⍴X
[22] YP←(P×YN)+((X-(P×XN))×B)
[23] MAPE←((+/((|(YP-YA))÷YA))÷N)×100
[24] RMSE←(((+/((YP-YA)*2))÷N)*0.5)÷((+/YA)÷N)
[25] YPA←(+/YP)÷N
[26] YAA←(+/YA)÷N
```

61

```
[27] YPM+YP-YPA
[28] YAM+YA-YAA
[29] NUM+(+/(YPM×YAM))+N
[30] DENOM+(((+/(YPM*2))+N)×((+/(YAM*2))+N))*0.5
[31] CORR+NUM+DENOM
[32] 'PREDICTED Y'
[33] YP
[34] '    '
[35] 'ACTUAL Y'
[36] YA
[37] '    '
[38] 'CORRELATION   ',⍕CORR
[39] '    '
[40] 'RMSE/MA      ',⍕RMSE
[41] '    '
[42] 'MAPE         ',⍕MAPE
     ∇


     ∇ PRED4;YPA;YAA;YPM;YAM;NUM;DENOM
[1] ⍝THIS FUNCTION COMPUTES THE MEASURES OF PREDICTIVE
[2] ⍝EFFECTIVENESS FOR THOSE CASES WHERE CORRECTIONS FOR
[3] ⍝FOURTH ORDER AUTOCORRELATION ARE MADE.
[4] 'ENTER A VECTOR CONTAINING THE ACTUAL Y VALUES.'
[5] OH2+⎕
[6] 'ENTER THE VECTOR OF X VALUES'
[7] DL2+⎕
[8] 'ENTER THE VALUE OF P4'
[9] P+⎕
[10] YA+⁻4↑OH2
[11] YN+⁻4↑OH1
[12] X+⁻4↑DL2
[13] XN+⁻4↑DL1
[14] N+ρX
[15] YP+(P×YN)+((X-(P×XN))×B)
```

62

```
[16] MAPE←((+/((|(YP-YA))÷YA))÷N)×100
[17] RMSE←(((+/((YP-YA)*2))÷N)*0.5)÷((+/YA)÷N)
[18] YPA←(+/YP)÷N
[19] YAA←(+/YA)÷N
[20] YPM←YP-YPA
[21] YAM←YA-YAA
[22] NUM←(+/(YPM×YAM))÷N
[23] DENOM←(((+/(YPM*2))÷N)×((+/(YAM*2))÷N))*0.5
[24] CORR←NUM÷DENOM
[25] 'PREDICTED Y'
[26] YP
[27] '     '
[28] 'ACTUAL Y'
[29] YA
[30] '     '
[31] 'CORRELATION   ',⍕CORR
[32] '     '
[33] 'RMSE/MA      ',⍕RMSE
[34] '     '
[35] 'MAPE        ',⍕MAPE
      ∇
```

```
      ∇ PRED40;YPA;YAA;YPM;YAM;NUM;DENOM
[1]  ⍝THIS FUNCTION COMPUTES THE MEASURES OF PREDICTIVE
[2]  ⍝EFFECTIVENESS FOR THOSE MODELS CORRECTING FOR
[3]  ⍝FIRST ORDER AND FOURTH ORDER AUTOCORRELATION:
[4]  'ENTER THE VECTOR CONTAINING ACTUAL Y VALUES'
[5]  OH2←□
[6]  'ENTER THE VECTOR CONTAINING X VALUES'
[7]  DL2←□
[8]  'ENTER A VALUE FOR P4'
[9]  P4←□
[10] 'ENTER THE VALUE FOR P1'
[11] P1←□
```

63

```
[12] 'ENTER THE VECTOR CONTAINING THE PREDICTED VALUES'
[13] YP←□
[14] YA←¯4↑OH2
[15] X←¯4↑DL2
[16] XT1←4↑(¯5↓DL2)
[17] XT4←4↑(¯8↓DL2)
[18] XT5←4↑(¯9↓DL2)
[19] XB←B×(((X-(P4×XT4))-(P1×XT1))+(P1×P4×XT5))
[20] YP←4ρ0
[21] YT1←5ρ0
[22] YT1[1]←1↑(¯5↓OH2)
[23] YT4←4↑(¯8↓OH2)
[24] YT5←4↑(¯9↓OH2)
[25] L←0
[26] ITERATE:L←L+1
[27] YP[L]←YT1[L+1]←(((P1×YT1[L])+(P4×YT4[L]))-(P1×P4×YT5[L]))+XB[L]
[28] →(L≤3)/ITERATE
[29] N←ρX
[30] MAPE←((+/((|(YP-YA))÷YA))÷N)×100
[31] RMSE←(((+/((YP-YA)*2))÷N)*0.5)÷((+/YA)÷N)
[32] YPA←(+/YP)÷N
[33] YAA←(+/YA)÷N
[34] YPM←YP-YPA
[35] YAM←YA-YAA
[36] NUM←(+/(YPM×YAM))÷N
[37] DENOM←(((+/(YPM*2))÷N)×((+/(YAM*2))÷N))*0.5
[38] CORR←NUM÷DENOM
[39] 'PREDICTED Y'
[40] YP
[41] '     '
[42] 'ACTUAL Y'
[43] YA
[44] '     '
[45] 'CORRELATION   ',⍕CORR
[46] '     '
```

64

```
[47] 'RMSE/MA      ',⍕RMSE
[48] '     '
[49] 'MAPE       ',⍕MAPE
     ∇
```

LIST OF REFERENCES

1. Boger, D. C., *Statistical Models For Estimating Overhead Costs*, Technical Report NPS-54-83-014, Department of Administrative Sciences, Naval Postgraduate School, 1983.

2. Maddala, G. S., *Econometrics*, McGraw-Hill Book Company, 1977.

3. O'Donovan, T. M., *Short Term Forecasting An Introduction to the Box-Jenkins Approach*, John Wiley and Sons, 1983.

4. Kmenta, J., *Elements of Econometrics, Second Edition*, Macmillan Publishing Company, 1986.

5. Park, R. E. and B. M. Mitchell, "Estimating the Autocorrelation Error Model with Trend Data," *Journal of Econometrics*, Vol. 13, No. 2, pp. 185-201, June 1980.

6. Wallis, K. F., "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations," *Econometrics*, Vol. 40, No. 4, pp. 617-636, July 1972.

7. Judge, G. C., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.C. Lee, *The Theory and Practice of Econometrics*, Second Edition, John Wiley and Sons, 1985.

8. Jerabek, D.J., "*Regression Models of Quarterly Overhead Costs for Six Government Aerospace Contractors*," March 1986, Masters Thesis, Naval Postgraduate School, 1986.

9. Box, G.E.P. and G.M. Jenkins, *Time Series Analysis Forecasting, Revised Edition, and Control*, Holden-Day, 1976.

10. Hoff, J.C., *A Practical Guide to Box-Jenkins Forecasting*, Lifetime Learning Publications, 1983.

11. Montgomery, D.C. and L.A. Johnson, *Forecasting and Time Series Analysis*, McGraw-Hill Book Company, 1976.

12. Tiao, G.C. and G.E.P. Box, "Modeling Multiple Time Series with Applications," *Journal of the American Statistical Association*, Vol. 76, No. 376, pp. 802-816, December 1981.

13. McCleary, R., Hay, R.A. Jr., Meidinger, E.E., and D. McDowall. *Applied Time Series Analysis for the Social Sciences*, Sage Publications, 1980.

14. Draper, N.R. and H. Smith. *Applied Regression Analysis, Second Edition*. John Wiley and Sons, 1981.

INITIAL DISTRIBUTION LIST

No. Copies

1. Defense Technical Information Center   2
   Cameron Station
   Alexandria, VA 22304-6145

2. Library, Code 0142   2
   Naval Postgraduate School
   Monterey, CA 93943-5002

3. Professor D. C. Boger   4
   Code 55Bc
   Naval Postgraduate School
   Monterey, CA 93943

4. CAPT D. Ramelli   2
   AIR-02B
   NAVAL AIR SYSTEM COMMAND
   Washington, D.C. 22202

5. LCDR C. P. Schnedar   2
   2007 S. Pennsylvania
   Roswell, NM 88201

END

Feb.

1988

DTIC